



UNIVERSITAS INDONESIA

**KLASIFIKASI DAN PENGUKURAN KUALITAS
SITASI MENGGUNAKAN *DEEP MULTI-TASK
LEARNING* DAN *MULTIPLE CRITERIA DECISION
MAKING***

RINGKASAN DISERTASI

**YANIASIH
1906341782**

**FAKULTAS ILMU KOMPUTER
PROGRAM DOKTOR ILMU KOMPUTER
DEPOK
2023**

ABSTRAK

Nama : Yaniasih
Pembimbing : Dr. Indra Budi, S.Kom, M.Kom

Analisis sitasi tradisional menghitung sama besar setiap sitasi dalam daftar pustaka. Metode ini dinilai kurang valid karena hanya melihat kuantitas bukan kualitas sitasi. Perkembangan kecerdasan buatan mendorong munculnya metode analisis sitasi dalam seluruh teks dokumen publikasi berbasis pada kualitas sitasi. Paradigma baru ini terus berkembang dan masih memiliki kekurangan dalam ketersediaan data, pengembangan metode, dan pemanfaatannya. Penelitian ini bertujuan untuk 1) mengembangkan metode *deep multi-task learning* (MTL) untuk klasifikasi tiga makna sitasi secara bersamaan, dan 2) menentukan bobot skor atribut sitasi dan kerangka pengukuran nilai kualitas sitasi dalam artikel. Penelitian menggunakan data sitasi jurnal Indonesia dari lima bidang ilmu yaitu pangan, kesehatan, sosial, energi, dan komputer. Set data terdiri dari 852 artikel dan 9.173 kalimat sitasi. Kalimat sitasi dianotasi manual untuk memperoleh data tiga makna sitasi yaitu sentimen, sumber peran, dan fungsi. Pada tahap ini disusun skema baru kategori fungsi sitasi. Model MTL dibangun untuk klasifikasi tiga makna tersebut secara bersamaan. Ada tiga jenis model MTL yang dikembangkan yaitu *shared-trunk*, *cross-stitched*, dan *shared-private* dengan menggunakan tiga arsitektur yaitu *convolutional neural network* (CNN), *long-short term memory* (LSTM), dan *bi-directional long-short term memory* (BiLSTM). Hasil klasifikasi model MTL dibandingkan dengan model mesin pembelajar tradisional dan *deep learning* tugas tunggal. Evaluasi kinerja menemukan model terbaik adalah MTL *shared-private* menggunakan CNN. Model ini memperoleh skor F1 makro 0,81 untuk klasifikasi sentimen, 0,87 untuk klasifikasi sumber peran, dan 0,90 untuk klasifikasi fungsi sitasi. Pengukuran kualitas sitasi dilakukan

menggunakan metode pembobotan *multiple criteria decision making* (MCDM). MCDM yang digunakan adalah *best worst method* untuk memperoleh bobot *ground truth*, dan dibandingkan dengan lima metode obyektif yaitu Standar Deviasi, Koefisien Gini, Entropi, *Criteria Importance Through Intercriteria Correlation* (CRITIC), dan *Method Based on the Removal Effects of Criteria* (MEREK). Hasil perbandingan bobot dan korelasi ranking Spearman menunjukkan metode MEREK menghasilkan bobot paling mendekati *ground truth*. Hasil ini memberikan bobot tertinggi sitasi pada bab pembahasan, sitasi bersumber peran dari data, sitasi bersentimen positif, dan sitasi berfungsi sebagai pembanding. Kerangka penilaian mengusulkan metrik Nilai Kualitas Sitasi (NKS) yang merupakan metrik baru dengan menjumlahkan bobot atribut lokasi, sentimen, sumber peran, dan fungsi sitasi per kemunculan (NKS_f), dalam satu dokumen (NKS_d), dan total dalam keseluruhan data (NKS_T).

Kata kunci: analisis sitasi, kualitas sitasi, *multi-task learning*, *multiple criteria decision making*, metrik sitasi

ABSTRACT

Traditional citation analysis counts each bibliography citation equally. It solely evaluates citation quantity, not quality. Artificial intelligence facilitates quality-based in-text citation analysis. Data availability, method, and implementation remain issues in this new paradigm. This paper aims to (1) develop a deep multi-task learning (MTL) model to classify three citation meanings simultaneously and (2) determine the weight of the citation attributes and the framework for measuring citation quality value. The data was the Indonesian journal article citations from five scientific fields: food, health, social, energy, and computer sciences. The dataset contained 852 articles and

9,173 citation sentences. Citation meanings, sentiment, role source, and function, were manually annotated. A new scheme for citation functions was designed at this level. MTL was developed to classify all three meanings at once. Three architectures, convolutional neural network (CNN), long-short term memory (LSTM), and bi-directional long-short memory (BiLSTM), were used to create three MTL models: shared-trunk, cross-stitched, and shared-private. Performance evaluation compared the MTL model to traditional machine learning and single-task deep learning models. The best model found was the MTL shared-private employing a convolutional neural network model. This model obtained a macro F1 score of 0.81 for sentiment classification, 0.87 for role source classification, and 0.90 for citation function classification. Measurement of citation quality was carried out using the multiple criteria decision-making (MCDM) weighting method. The MCDM used was the best worst method to obtain ground truth weights, and then compared to five objective methods, namely Standard Deviation, Gini Coefficient, Entropy, Criteria Importance Through Intercriteria Correlation (CRITIC), and Method Based on the Removal Effects of Criteria (MERECE). The weight comparison and Spearman ranking correlation results showed that the MERECE method produced the weight closest to the ground truth. These results gave citations the highest weight in the discussion chapter, citations originate from the role of data, citations have positive sentiments, and citations serve as a comparison. The assessment framework proposed a Citation Quality Score (NKS) metric, a new metric that sums up the weights of location, sentiments, role sources, and functions of a single citation occurrence in the text (NKS_f), all same citation in one citing document (NKS_d), and total same citation in set data (NKS_T).

Keywords: citation analysis, citation quality, *multi-task* learning, multiple criteria decisions making, citation metrics

1. PENDAHULUAN

1.1. Latar belakang

Riset analisis sitasi mengkaji jumlah, pola, dan jaringan dari artikel yang disitir dalam dokumen yang menyitir. Analisis sitasi muncul dari asumsi bahwa sitasi dapat memberikan informasi mengenai hubungan antar artikel, sejarah perkembangan ide, dan penemuan topik riset (De Bellis, 2010). Asumsi kedua yang berkembang luas adalah pemanfaatannya untuk melakukan evaluasi dan pemeringkatan sains (Gingras, 2014). Hal ini menimbulkan pro dan kontra karena dihubungkan dengan penilaian promosi dan jabatan. Di beberapa negara digunakan sebagai indikator dalam memberikan pendanaan riset serta penghargaan bagi individu dan lembaga (White, 2004), termasuk di Indonesia.

Sitasi yang selama ini umum dianalisis adalah jumlah sitasi dalam daftar pustaka. Metode ini dinilai kurang valid karena hanya mengukur kuantitas bukan kualitas sitasi (Maricic et al., 1998; Shahid et al., 2015). Penelitian pendahuluan sudah dilakukan untuk menganalisis kekurangan analisis sitasi tradisional. Analisis teori kritis filsafat ilmu pengetahuan memperlihatkan permasalahan analisis sitasi tradisional terletak pada konsep, metode, dan implementasi. Kekurangan dimulai dari paradigma positivisme yang gagal mencapai tujuan untuk menjamin obyektivitas dan bebas nilai dalam pencarian kebenaran. Dari sisi metode, kritik ditekankan pada bias sumber data, teknis pengumpulan data, metode analisis, dan interpretasi hasil yang sangat tergantung pada kemampuan teknis peneliti. Kritik yang ketiga terletak pada isu bias ideologi dan etika dalam implementasi khususnya penggunaan analisis sitasi untuk evaluasi kinerja riset (Yaniasih, 2020).

Kajian teori kritis menunjukkan banyak kekurangan dalam konsep dan metode analisis sitasi berbasis daftar pustaka. Perspektif ilmu komputer memunculkan solusi ke arah analisis sitasi dalam seluruh bagian teks. Konsep analisis sitasi dalam

teks dapat membedakan makna dalam sitasi berdasarkan lokasi, sentimen, dan tujuan penulis dalam mengutip. Metode analisis sitasi dalam teks menjadi rekomendasi untuk memperbaiki dan meningkatkan pemanfaatan analisis sitasi menjadi lebih baik. Gambar 1.1. menampilkan komponen dan tujuan analisis sitasi dalam teks yang dilakukan dalam penelitian ini.



Mengukur kualitas sitasi yaitu tingkat kepentingan sitasi (artikel yang disitir) dalam artikel yang menyitir

Analisis sitasi dalam teks:
 Sitasi: Strydom, 2011
Intensitas: frekuensi muncul = 1
Lokasi: bab metode
Konteks: kalimat sitasi "kajian ini mengadopsi teori kritis kontemporer" diklasifikasi:
 → Sentimen = netral
 → Fungsi = menggunakan
 → Sumber peran = metode

Sitasi dalam teks artikel jurnal

Referensi dalam daftar pustaka → menghitung jumlah sitasi

Gambar 1.1. Komponen dan tujuan analisis sitasi dalam teks

1.2. Perumusan masalah

1) Bagaimana metode MTL terbaik untuk klasifikasi tiga makna kalimat sitasi dalam teks secara bersamaan?

Klasifikasi tiga makna sitasi menggunakan data yang sama dan saling berkaitan. Model MTL diharapkan dapat mengklasifikasi tiga makna secara bersamaan dan memberikan hasil klasifikasi lebih baik dibandingkan model tugas tunggal. Ada beberapa permasalahan dalam klasifikasi tiga makna sitasi secara bersamaan antara lain:

- Upaya untuk mengatasi data yang tidak seimbang seperti *resampling* pada set data klasifikasi menggunakan MTL menjadi rumit karena data terdiri dari satu input untuk tiga target output. Sementara teknik *resampling* hanya berlaku untuk 1 input dengan 1 target output sehingga tidak bisa digunakan untuk data klasifikasi MTL. Ketidak-seimbangan data menjadi tantangan karena pada kelas sentimen dan sumber peran perbedaan proporsi data cukup besar.
 - Tantangan optimisasi *hyper-parameter* terkait penyusunan model karena metode optimisasi yang ada secara teknis tidak mudah untuk diaplikasikan pada model multi-output seperti MTL. Hal ini disebabkan secara umum optimisasi berdasarkan pada nilai *loss* tunggal dan *library* yang umum dipakai tidak mendukung optimisasi MTL.
- 2) Bagaimana bobot skor dan kerangka pengukuran yang tepat untuk penilaian kualitas sitasi menggunakan atribut sitasi dalam teks?
- Permasalahan dalam pembobotan adalah diperlukan data standar atau pembanding untuk evaluasi bobot hasil eksperimen. Belum ditemukan bobot atribut sitasi secara terperinci untuk semua atribut sitasi dalam teks dan sebagian besar usulan bobot dari literatur sebelumnya adalah berdasarkan opini atau pendapat subyektif penulis. Oleh karena itu diperlukan bobot sebagai kebenaran dasar (*ground truth*) dan pemilihan metode yang dapat menghasilkan bobot mendekati *ground truth*.
 - Bobot yang dihasilkan kemudian digunakan dalam kerangka pengukuran nilai kualitas sitasi. Belum ada referensi yang dapat dirujuk untuk membangun kerangka pengukuran kualitas sitasi menggunakan atribut lokasi dan tiga makna sitasi secara bersamaan.

1.3. Tujuan

- 1) Mengembangkan metode MTL untuk klasifikasi tiga makna kalimat sitasi secara bersamaan.
- 2) Menentukan bobot skor dari atribut sitasi dan kerangka pengukuran nilai kualitas sitasi dalam artikel jurnal.

1.4. Keluaran penelitian

- 1) Model MTL untuk klasifikasi tiga makna sitasi
Penelitian mengembangkan model MTL untuk klasifikasi tiga makna sitasi secara bersamaan. Pengembangan metode berkontribusi melengkapi teknik klasifikasi konteks kalimat sitasi dalam teks artikel jurnal.
- 2) Bobot skor atribut sitasi dan kerangka penilaian kualitas sitasi
Hasil bobot skor atribut sitasi dan kerangka penilaian kualitas sitasi melengkapi literatur bidang kajian kuantitatif sains, ilmu informasi, dan ilmu komputer, terutama berkaitan dengan topik perolehan informasi ilmiah dan evaluasi kinerja riset.

1.5. Manfaat penelitian

- 1) Informasi makna sentimen, sumber peran, dan fungsi dari kalimat konteks sitasi dapat digunakan dalam pencarian sumber rujukan yang tepat oleh mahasiswa dan peneliti dalam melaksanakan riset dan penulisan artikel ilmiah.
 - Makna sentimen yang memberitahukan pendapat penulis lain tentang kelebihan atau kekurangan suatu artikel dapat menjadi pertimbangan ketika akan menyitir suatu artikel.
 - Makna sumber peran dan fungsi dapat mempermudah pencarian literatur rujukan dengan tujuan khusus. Sebagai contoh ketika mencari/membutuhkan rujukan untuk membahas hasil penelitian maka dapat dipilih artikel yang disitir karena sumber perannya hasil dan fungsinya pembanding.

- 2) Informasi sitasi untuk setiap artikel dapat diperkaya terdiri dari:
- jumlah sitasi berdasarkan jumlah artikel yang menyitir (berdasarkan referensi di daftar pustaka)
 - jumlah sitasi dalam teks yang memperlihatkan penggunaan sesungguhnya referensi dalam penulisan artikel
 - nilai kualitas sitasi berdasarkan lokasi dan tiga makna sitasi

Informasi yang lebih lengkap tersebut dapat digunakan oleh individu/penerbit/lembaga untuk mengevaluasi kualitas publikasi mereka. Lebih lanjut, hasil evaluasi dapat digunakan dalam pengambilan keputusan dan kebijakan riset di skala lembaga maupun nasional.

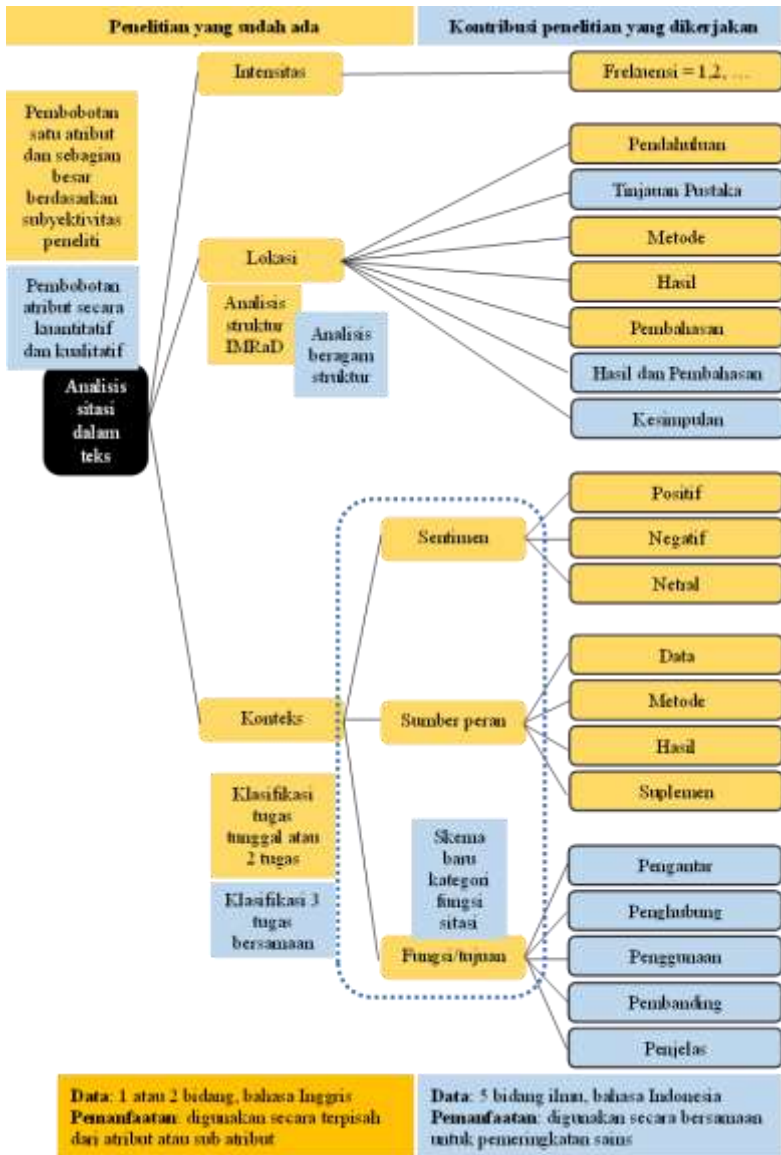
BAB 2. STUDI PUSTAKA

Studi pustaka menggunakan teknik bibliometrik dan kajian literatur sistematis. Sumber literatur berasal dari basis data Scopus (<https://scopus.com>) dan Google Scholar/GS (<https://scholar.google.com/>). Hasil pencarian dan penyaringan pustaka memperoleh 248 dokumen untuk analisis bibliometrik dan 52 dokumen untuk *review* secara sistematis.

Analisis bibliometrik terhadap pustaka sitasi dalam teks memperlihatkan topik ini masih terus berkembang. Perspektif dari negara berkembang sangat dibutuhkan terlihat dari analisis penulis dan sumber jurnal yang sebagian besar dari negara maju. Tren kata kunci menggambarkan adanya evolusi dari analisis sitasi deskriptif ke arah analisis makna menggunakan teknologi pembelajaran mesin. Analisis *co-citation* berhasil mengidentifikasi literatur utama dalam konsep dan metode analisis sitasi dalam teks. Analisis terhadap literatur utama memberikan gambaran perkembangan ilmu dan teknologi analisis sitasi dalam teks. Klasifikasi untuk menentukan tingkat

kepentingan sitasi menjadi tren penelitian terkini terutama klasifikasi makna kalimat sitasi.

Analisis lebih mendalam terhadap konten pustaka menunjukkan masih ada gap dan kekurangan dalam penelitian yang sudah ada. Kekurangan pada data yang digunakan masih terbatas jumlahnya, cakupan bidang ilmu, dan bahasa kalimat yang dianalisis. Kekurangan metode ditemukan setidaknya pada tiga hal yaitu (1) atribut lokasi yang dianalisis hampir sebagian besar merujuk pada struktur IMRaD padahal banyak jurnal yang strukturnya tidak standar IMRaD, (2) klasifikasi makna sitasi baru terbatas pada satu dan dua makna, sedangkan makna sitasi yang menggunakan data sama lebih dari dua dan memungkinkan diklasifikasi secara bersamaan, dan (3) belum ditemukan metode dan usulan bobot skor semua atribut sitasi. Kekurangan pada implementasi terletak pada belum adanya penelitian menggunakan semua atribut sitasi untuk pemeringkatan sains, apalagi diketahui salah satu pemanfaatan utama analisis sitasi adalah untuk evaluasi sains. Ilustrasi lebih jelas mengenai kekurangan penelitian yang sudah ada dan kontribusi penelitian yang dikerjakan dipetakan dalam Gambar 2.1. Ringkasan kebaruan penelitian ini disajikan pada Tabel 2.1.



Gambar 2.1. Pemetaan penelitian analisis sitasi dalam teks

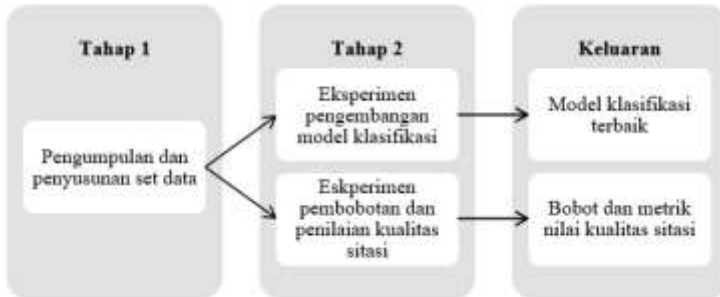
Tabel 2.1. Ringkasan kebaruan penelitian

Aspek	Perkembangan literatur	Kebaruan penelitian
Data	<ul style="list-style-type: none">- Artikel jurnal International- Diterbitkan di negara maju- Terbatas pada satu atau dua bidang ilmu tertentu	Melengkapi literatur dari data, perspektif, dan pemanfaatan praktis yang berbeda: <ul style="list-style-type: none">- Jurnal Bahasa Indonesia- Terbit di negara berkembang- Multi bidang : pangan, kesehatan, sosial, energi, dan komputer
Metode	<ul style="list-style-type: none">- Klasifikasi manual- <i>Rule-based, traditional ML, single-output DL, multi-task DL</i> untuk dua tugas- Konsep umum penilaian kualitas sitasi	<ul style="list-style-type: none">- Mengembangkan model <i>deep multi-task learning</i> untuk tiga tugas klasifikasi- Menentukan bobot skor, nilai, dan ranking kualitas sitasi
Analisis/ Pemanfaatan	<ul style="list-style-type: none">- Satu atau dua makna- Sistem rekomendasi sitasi, peringkasan dokumen	<ul style="list-style-type: none">- Analisis tiga makna sitasi bersamaan- Pemingkatan sains

3. METODOLOGI

Penelitian terbagi menjadi dua tahap. Tahap pertama adalah pengumpulan dan penyusunan data untuk dianalisis. Hasil tahap satu adalah set data sitasi dalam teks. Tahap kedua terdiri atas dua bagian yaitu (1) eksperimen pengembangan model klasifikasi, (2) pembobotan dan pengembangan kerangka penilaian kualitas sitasi. Dua bagian ini dapat dilakukan secara paralel untuk menjawab pertanyaan penelitian satu dan dua. Hasil tahap dua kemudian digunakan untuk

membangun prototipe implementasi. Kerangka tahapan eksperimen penelitian secara umum disajikan pada Gambar 3.1.

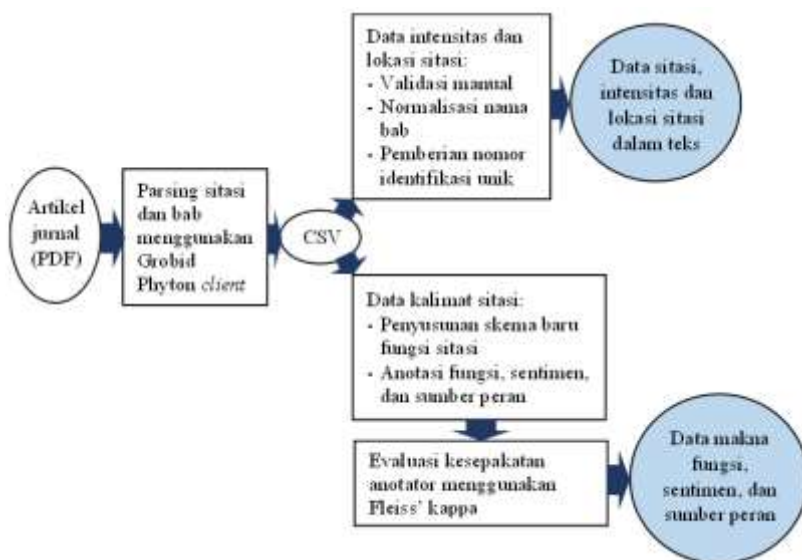


Gambar 3.1. Kerangka tahapan eksperimen penelitian

3.1. Pengumpulan dan penyusunan data eksperimen

Tahapan penyusunan set data disajikan dalam Gambar 3.2. Data yang digunakan adalah sitasi dalam teks artikel jurnal Indonesia yang masuk dalam peringkat 1 dan 2 basis data SINTA Kemenristek/BRIN. Jumlah jurnal yang dikumpulkan sebanyak 29 jurnal, terdiri atas 852 artikel dan 9.173 kalimat.

Artikel jurnal terpilih diunduh dalam file PDF, kemudian diurai (*parsing*) referensi dan lokasi sitasi dalam teks menggunakan Grobid (Lopez, 2009) berbasis Phyton client. Hasil parsing adalah ekstraksi referensi, sitasi dalam teks, lokasi sitasi, dan kalimat dalam format XML. Data dalam format XML kemudian diubah ke format CSV dan tabel variabel sitasi dalam teks dibuat. Variabel intensitas dan lokasi sitasi divalidasi manual. Normalisasi nama bab dilakukan sesuai format IMRaD dan modifikasinya. Pencocokan sitasi dilakukan dengan memberikan nomor identifikasi unik karena belum semua sitasi memiliki nomer DOI (*Digital Object Identifier*).



Gambar 3.2. Tahapan penyusunan set data sitasi dalam teks

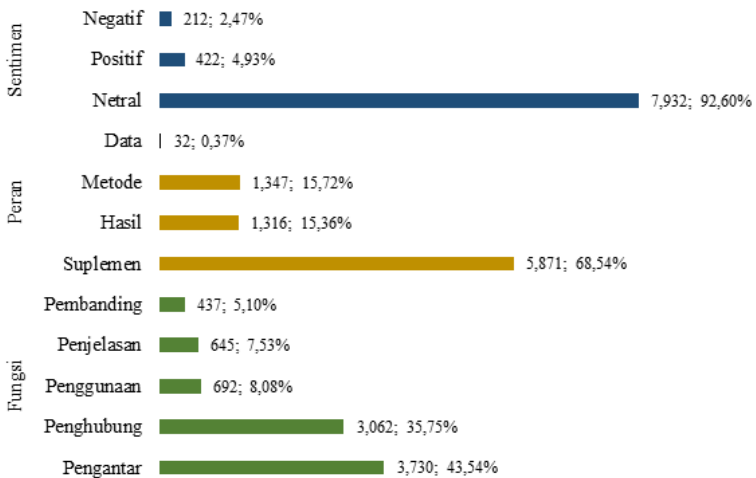
Kalimat konteks sitasi hasil *parsing* kemudian dinotasi secara manual. Tiga orang ahli memberikan label untuk setiap kalimat. Skema kategori label mengacu pada *state of the art*. Kategori sentimen terdiri dari positif, negatif, dan netral. Makna sumber peran mengacu pada Zhao et al (2019). Khusus makna fungsi sitasi, karena sudah banyak penelitian yang dilakukan dengan skema yang berbeda-beda, maka dilakukan penelitian pendahuluan untuk menyusun skema secara sistematis sesuai data dan tujuan penelitian ini. Penyusunan skema ini sudah dipublikasikan di jurnal Publications (Yaniasih dan Budi, 2021). Hasil penyusunan tersebut menghasilkan data yang sangat tidak seimbang. Oleh karena itu, tanpa mengubah makna dari skema awal yang disusun, penelitian ini memodifikasi skema tersebut. Skema label kategori untuk tiga makna ditampilkan pada Tabel 3.1. Adapun data hasil anotasi disajikan pada Gambar 3.3.

Tabel 3.1. Skema label kategori makna sitasi

Atribut	Kategori	Definisi
Sentimen sitasi	Positif	Sentimen masuk kategori positif ketika kalimat sitasi menunjukkan pujian, kelebihan, konfirmasi, manfaat, dan kecenderungan positif lainnya.
	Negatif	Sentimen negatif ditunjukkan ketika kalimat menyampaikan kekurangan, kelemahan, kontras, ketidaksesuaian, perlunya perbaikan, dan indikasi negatif lainnya.
	Netral	Sentimen netral ketika kalimat sitasi tidak menunjukkan kecenderungan.
Sumber peran sitasi	Data	Kalimat konteks sitasi menunjukkan artikel disitir berupa data atau material yang digunakan kembali dalam penelitian/tulisan yang menyitir
	Metode	Kalimat konteks sitasi menunjukkan artikel disitir berupa metode, prosedur, alat (<i>tool</i>), kode, algoritma, atau suatu cara yang digunakan, diadopsi, dimodifikasi, atau dilanjutkan dalam penelitian/tulisan yang menyitir
	Hasil	Kalimat konteks sitasi menunjukkan artikel disitir merupakan hasil dari suatu penelitian
	Suplemen	Kalimat konteks sitasi menunjukkan artikel disitir adalah selain data, metode, dan hasil yaitu berupa pernyataan, berita, media, dll.
Fungsi sitasi	Pemanfaatan (<i>as basic, used, and extend</i>)	Merupakan gabungan dari penggunaan dan kelanjutan. Artikel yang disitir digunakan secara langsung dalam riset, seperti penggunaan konsep, metode, data, rumus, algoritma, dll, atau artikel yang dijadikan dasar dan dilanjutkan dalam penelitian yang ditulis. Umumnya ditandai dengan kata menggunakan, digunakan, berdasarkan, melanjutkan, mengembangkan, mengikuti, mengadopsi, memodifikasi, dimotivasi, dst.

Tabel 3.1. Lanjutan

Pembandingan	Artikel yang disitir dijadikan pembandingan (terutama hasil) bagi artikel yang ditulis. Ditandai dengan kata dibandingkan, sesuai, mirip, kesamaan, konfirmasi, berbeda, kelemahan, keunggulan, kontras, sejalan, bertentangan, lebih tinggi, lebih rendah, serupa, dst.
Penghubung	Artikel yang disitir memiliki keterkaitan atau membahas secara khusus/rinci bahan, metode, proses, atau data hasil yang terkait langsung. Kata yang menandai antara lain penelitian (-sebelumnya, -lain), terkait, sedangkan, telah dilakukan, melaporkan, mendefinisikan, menurut, literatur, dst.
Pengantar	Artikel yang disitir memberikan informasi atau teori umum (tidak spesifik bidang ilmu) untuk mengenalkan, mencontohkan, menjadi latar belakang, masalah, tujuan, dst.
Penjelas	Artikel yang disitir menjelaskan hasil atau pernyataan dalam tulisan.



Gambar 3.3. Data label per makna sitasi

3.2. Pengembangan model dan eksperimen klasifikasi

Tahap pengembangan model terdiri dari penyusunan model dasar, optimasi *hyperparameter*, dan evaluasi model optimasi. Model dasar menggunakan arsitektur CNN, LSTM, dan BiLSTM merujuk pada model dasar yang digunakan dalam penelitian-penelitian sebelumnya (Crawshaw, 2020; Liu et al., 2017; Misra et al., 2016; Ruder, 2017; Zhang dan Yang, 2021). Terdapat tiga jenis arsitektur MTL yang digunakan yaitu (1) *shared trunk* (ST), *cross-stitched* (CS), dan MTL *shared-private* (SP). Proses optimasi *hyperparameter* menggunakan perangkat lunak Optuna dengan metode sampling *Tree-structured Parzen Estimators* (TPE). *Hyperparameter* yang dioptimasi meliputi nilai *embedding*, *filter*, *kernel*, unit CNN/LSTM/Bi-LSTM, *dense*, *dropout*, tingkat pembelajaran (*learning rate*), dan *batch*. Daftar *hyperparameter* dan alternatif nilai disajikan dalam Tabel 3.2.

Tabel 3.2. Optimasi *hyper-parameter*

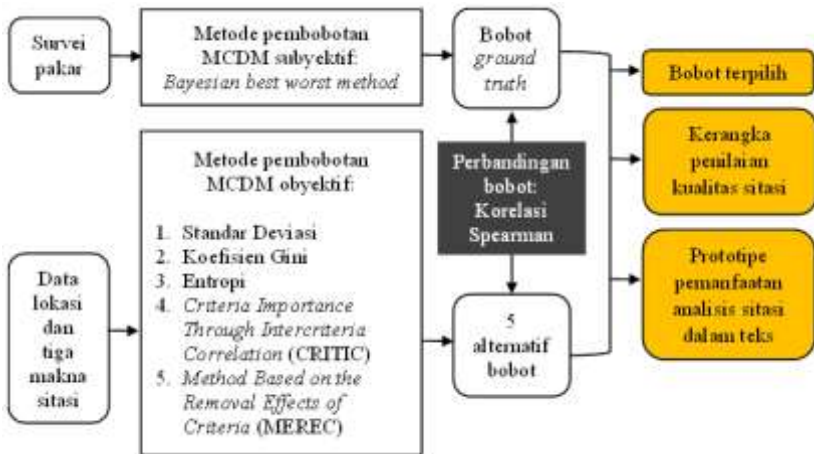
<i>Hyper-parameter</i>	Alternatif nilai
Embedding	32, 64, 128
Filter CNN	32, 64, 128
Kernel CNN	3, 4, 5
Unit LSTM/BiLSTM	32, 64, 128
Dense	32, 64, 128
Dropout	0, 0,25, 0,5, 0,75
Learning rate	1, 0,1, 0,01, 0,001, 0,0001
Batch	16, 32, 64

Model yang dioptimalkan kemudian dievaluasi kinerjanya dan dibandingkan dengan beberapa metode yang digunakan dalam penelitian sebelumnya. Model pembanding adalah Nave Bayes (NB), Random Forrest (RF), Support Vector Machine (SVM), Long-short Term Memory (LSTM), dan model Bidirectional (Bi-LSTM). Proses pelatihan dan validasi mesin menggunakan validasi silang. Kemampuan klasifikasi

diukur dengan menggunakan metrik akurasi, presisi, daya ingat, dan skor F1 makro (Lever et al., 2016).

3.3. Pembobotan atribut dan kerangka penilaian kualitas sitasi

Penentuan bobot skor menggunakan enam metode pembobotan *multiple criteria decision making* (MCDM). Sampai saat ini belum ditemukan literatur bobot skor atribut sitasi secara lengkap sehingga diperlukan bobot yang dijadikan kebenaran dasar (*ground truth*). Tahapan eksperimen penentuan bobot ditampilkan pada Gambar 3.4.



Gambar 3.4. Tahapan eskperimen penentuan bobot atribut

Bobot *ground truth* dalam penelitian ini dihasilkan dari hasil survei grup pakar menggunakan metode MCDM subyektif yaitu *best worst method* (BWM). Hasil bobot BWM kemudian dibandingkan dengan bobot hasil lima metode obyektif MCDM. Perbandingan ini dilakukan untuk memperoleh metode otomatis yang tidak tergantung pada ketersediaan pakar dalam menentukan bobot atribut sitasi sehingga dapat digunakan pada penelitian-penelitian selanjutnya.

Bobot terpilih digunakan untuk menyusun kriteria dan rumus penilaian kualitas sitasi. Rumus metrik yang dihasilkan merupakan pendekatan baru dalam menilai sitasi sehingga perlu diuji validitasnya. Evaluasi yang digunakan dalam penelitian ini adalah membandingkan ranking hasil NKS dengan ranking sitasi yang dilakukan oleh ahli. Dua orang ahli bidang ilmu komputer menilai kualitas sitasi dari sampel sepuluh artikel. Hasil perbandingan diukur menggunakan korelasi Spearman.

4. EKSPERIMEN KLASIFIKASI MAKNA SITASI

4.1. Eskperimen klasifikasi sentimen sitasi

Eskperimen klasifikasi sentimen sitasi ditampilkan pada Tabel 4.1. Hasil klasifikasi sentimen memperoleh akurasi $\geq 0,90$ pada semua model. Nilai akurasi tertinggi didapatkan oleh model *multi-task* share-private model (MTL PL) CNN dengan nilai 0,97, sedangkan akurasi terendah diperoleh oleh model *single-task* RF menggunakan fitur n-gram vector. Nilai presisi pada semua model sangat bervariasi. Model NB memperoleh nilai presisi terendah (0,49) sedangkan model RF dan MTL SP CNN memperoleh nilai tertinggi masing-masing 0,89. Adapun nilai *recall* pada sebagian besar model sangat rendah ($<0,60$) kecuali pada tiga model yaitu *single-task* CNN, MTL CS CNN, dan MTL SP CNN yang memperoleh nilai $>0,60$. *Single-task* CNN mendapatkan nilai *recall* tertinggi yaitu 0,87. Nilai *recall* yang rendah menyebabkan nilai F1 juga rendah pada sebagian besar model ($\leq 0,70$), sehingga hanya tiga model dengan nilai *recall* tinggi yang memperoleh nilai F1 tinggi. Tiga model dengan nilai F1 tertinggi berturut-turut diperoleh oleh model *single-task* CNN dengan nilai 0,85, model MTL SP CNN dengan nilai 0,81, dan model MTL CS CNN dengan nilai 0,75.

Tabel 4.1. Perbandingan kinerja model klasifikasi sentimen

Metode	A	P	R	F	Waktu
Single task NB+n-gram vector	0,90	0,49	0,45	0,47	58d
Single task LR+n-gram vector	0,94	0,71	0,59	0,63	1m4d
Single task SVM+n-gram TF-IDF	0,93	0,72	0,58	0,63	59d
Single task RF+n-gram vector	0,94	0,89	0,51	0,59	55d
Single task LSTM	0,94	0,53	0,48	0,50	1j4m54d
Single task BiLSTM	0,95	0,77	0,53	0,57	1j24m11d
Single task CNN	0,91	0,86	0,87	0,85	25m8d
<i>Multi-task</i> FS LSTM	0,94	0,52	0,46	0,48	30m39d
<i>Multi-task</i> FS BiLSTM	0,94	0,52	0,50	0,51	30m53d
<i>Multi-task</i> FS CNN	0,94	0,65	0,47	0,49	3m29d
<i>Multi-task</i> CS LSTM	0,94	0,64	0,47	0,51	27m24d
<i>Multi-task</i> CS BiLSTM	0,95	0,68	0,55	0,60	3j33m48d
<i>Multi-task</i> CS CNN	0,96	0,85	0,69	0,75	1j7m39d
<i>Multi-task</i> SP LSTM	0,94	0,64	0,49	0,53	40m24d
<i>Multi-task</i> SP BiLSTM	0,95	0,85	0,55	0,57	1j10m19d
<i>Multi-task</i> SP CNN	0,97	0,89	0,75	0,81	37m11d

* A = Akurasi, P= Presisi, R = *Recall*, F = skor F1 makro

4.2. Eskperimen klasifikasi sumber peran sitasi

Eskperimen klasifikasi sumber peran sitasi ditampilkan pada Tabel 4.2. Klasifikasi peran menggunakan set data yang lebih seimbang dan memperoleh kinerja yang lebih baik daripada klasifikasi sentimen. Nilai akurasi antara 0,77-0,97 lebih rendah dibanding rata-rata nilai akurasi klasifikasi sentimen. Namun nilai presisi pada semua model $\geq 0,70$ dan pada *recall* $\geq 0,64$ jauh lebih baik dan merata dibandingkan klasifikasi sentimen. Adapun nilai F1 pada semua model adalah $\geq 0,70$ dan nilai tertinggi diperoleh oleh model MTL SP CNN dengan nilai 0,87. Model ini juga memperoleh nilai *recall* tertinggi (0,84) dan nilai presisi yang baik (0,93) walaupun bukan yang paling tinggi. Model lain yang memiliki nilai F1 $> 0,80$ adalah MTL SP CNN dan MTL CS CNN (0,86); MTL FS CNN (0,84); LR, SVM, dan RF (0,83); dan *single-task* CNN (0,81).

Tabel 4.2. Perbandingan kinerja model klasifikasi sumber peran

Metode	A	P	R	F	Waktu
Single task NB+n-gram vector	0,77	0,78	0,77	0,77	2m22d
Single task LR+n-gram vector	0,83	0,83	0,83	0,83	2m11d
Single task SVM+n-gram TF-IDF	0,83	0,83	0,83	0,83	1m53d
Single task RF+n-gram vector	0,83	0,83	0,83	0,83	2m40d
Single task LSTM	0,96	0,70	0,71	0,71	1j4m54d
Single task BiLSTM	0,97	0,71	0,72	0,71	1j24m11d
Single task CNN	0,95	0,92	0,77	0,81	25m8d
<i>Multi-task</i> FS LSTM	0,91	0,66	0,64	0,65	30m39d
<i>Multi-task</i> FS BiLSTM	0,90	0,66	0,64	0,65	30m53d
<i>Multi-task</i> FS CNN	0,96	0,94	0,77	0,84	3m29d
<i>Multi-task</i> CS LSTM	0,95	0,70	0,70	0,70	27m24d
<i>Multi-task</i> CS BiLSTM	0,96	0,97	0,74	0,77	3j33m33d
<i>Multi-task</i> CS CNN	0,97	0,97	0,82	0,86	1j7m39d
<i>Multi-task</i> SP LSTM	0,95	0,94	0,71	0,72	40m24d
<i>Multi-task</i> SP BiLSTM	0,96	0,93	0,82	0,86	30m53d
<i>Multi-task</i> SP CNN	0,94	0,93	0,84	0,87	37m11d

* A = Akurasi, P= Presisi, R = *Recall*, F = skor F1 makro

4.3. Eskperimen klasifikasi fungsi sitasi

Eskperimen klasifikasi fungsi sitasi ditampilkan pada Tabel 4.3. Hasil klasifikasi fungsi sitasi memperlihatkan model-model DL memiliki kinerja lebih baik dari semua model mesin pembelajaran tradisional (ML). Semua model ML memperoleh akurasi, presisi, *recall*, dan F1 $\leq 0,65$. Adapun model-model DL memperoleh akurasi terendah 0,86 dan tertinggi 0,93, presisi terendah 0,83 dan tertinggi 0,92, *recall* terendah 0,78 dan tertinggi 0,90, serta F1 terendah 0,80 dan tertinggi 0,90. Model dengan nilai F1 tertinggi adalah MTL SP CNN. Model ini juga memiliki skor tertinggi untuk nilai *recall*.

Tabel 4.3. Perbandingan kinerja model klasifikasi fungsi

Metode	A	P	R	F	Waktu
Single task NB+n-gram vector	0,60	0,52	0,48	0,50	6m38d
Single task LR+n-gram vector	0,64	0,61	0,55	0,58	7m
Single task SVM+n-gram TF-IDF	0,63	0,59	0,55	0,57	6m47d
Single task RF+n-gram vector	0,65	0,72	0,51	0,54	6m35d
Single task LSTM	0,90	0,83	0,83	0,83	1j4m54d
Single task BiLSTM	0,93	0,89	0,82	0,85	1j24m11d
Single task CNN	0,88	0,83	0,88	0,89	25m8d
<i>Multi-task</i> FS LSTM	0,86	0,87	0,80	0,82	30m39d
<i>Multi-task</i> FS BiLSTM	0,92	0,92	0,87	0,87	30m53d
<i>Multi-task</i> FS CNN	0,91	0,87	0,89	0,88	3m29d
<i>Multi-task</i> CS LSTM	0,90	0,90	0,88	0,89	27m24d
<i>Multi-task</i> CS BiLSTM	0,91	0,90	0,86	0,88	3j33m48d
<i>Multi-task</i> CS CNN	0,88	0,83	0,78	0,80	1j7m39d
<i>Multi-task</i> SP LSTM	0,92	0,87	0,78	0,80	40m24d
<i>Multi-task</i> SP BiLSTM	0,90	0,92	0,87	0,87	30m53d
<i>Multi-task</i> SP CNN	0,92	0,91	0,90	0,90	37m11d

* A = Akurasi, P= Presisi, R = *Recall*, F = skor F1 makro

4.4. Pemilihan model terbaik

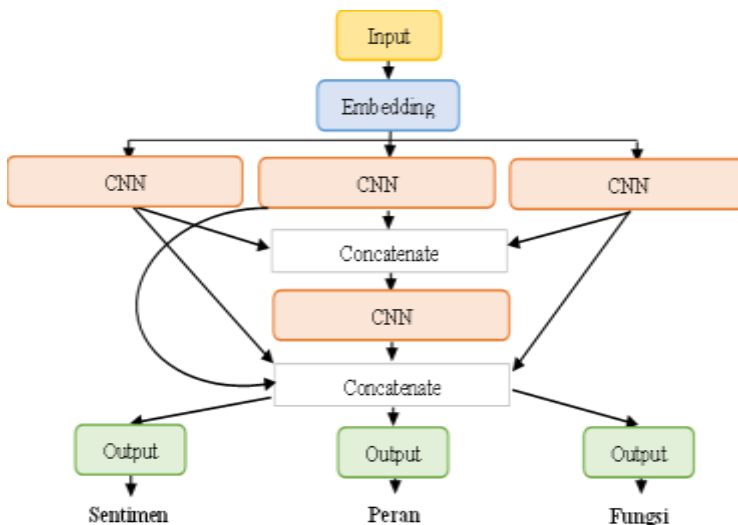
Salah satu tujuan dari model *multi-output* adalah efisiensi. Efisiensi waktu pelatihan pada setiap model menunjukkan tidak ada pola yang sama antar model dan arsitektur. Secara umum dari sisi waktu, model yang paling efisien adalah CNN menggunakan MTL FS. Di posisi kedua ada lima model yang memiliki waktu hampir sama yaitu *single-task* ML, *single-task* CNN, MTL FS dan SP BiLSTM, serta MTL SP CNN.

Selain efisiensi waktu, dalam pengembangan model perlu diperhatikan kompleksitas dari algoritma. Model MTL lebih kompleks dibandingkan model *single-task*. Antar model MTL, model FS adalah yang paling sederhana, sedangkan model SP adalah yang paling kompleks. Namun, perbedaan kompleksitas antar ketiga model tersebut tidak terlalu besar karena konsep dan susunanya hampir mirip.

Walaupun efisiensi menjadi pertimbangan, namun dasar utama pemilihan model adalah pada kinerjanya. Apalagi ketika efisiensi tidak terlalu jauh berbeda antar model. Pembahasan hasil klasifikasi pada tiga makna sitasi menunjukkan model CNN baik *single-* maupun *multi-task* memiliki kinerja lebih baik dibandingkan model ML maupun DL dengan arsitektur LSTM dan BiLSTM. Model *single-task* CNN memperoleh kinerja terbaik pada klasifikasi sentimen, terbaik kedua pada klasifikasi fungsi, namun untuk klasifikasi peran memiliki kinerja jauh di bawah model-model lain pada klasifikasi peran. Model MTL SP CNN memperoleh nilai F1 tertinggi pada klasifikasi peran dan fungsi, namun pada klasifikasi sentimen tidak sebaik model *single-task* CNN. Merujuk pada tujuan penelitian untuk membangun model yang dapat mengklasifikasi tiga makna sitasi secara bersamaan, model MTL SP CNN menjadi kandidat model terbaik.

Analisis lebih mendalam untuk menentukan model terbaik dilakukan pada kinerja klasifikasi per kelas makna sitasi. Karena jumlah kelas yang tidak seimbang, perhatian diberikan pada kemampuan model untuk mengklasifikasikan kelas minoritas. Model MTL SP CNN berhasil mengklasifikasikan semua kelas dengan baik, termasuk kelas minoritas, dengan nilai setiap metrik >50%. Kelas positif pada makna sentimen dengan jumlah data 4,93% memperoleh nilai 0,71. Adapun kelas negatif yang jumlahnya sangat kecil (2,47%) berhasil diklasifikasi dengan nilai F1 0,66. Nilai *recall* yang sering menjadi metrik dengan nilai terendah memperoleh nilai 0,54. Adapun pada makna peran, kelas yang paling kecil jumlahnya, kurang dari 1% (0,37%) adalah kelas data. Model MTL SP CNN mampu mengklasifikasi kelas ini dengan nilai F1 0,79. Adapun kelas dengan jumlah data yang besar, seperti kategori netral dan suplemen, serta semua kategori dalam klasifikasi fungsi, memperoleh nilai F1 di atas 80%. Berdasarkan analisis terhadap metrik kinerja per makna dan per kelas, model MTL SP CNN merupakan model dengan kinerja

terbaik untuk klasifikasi tiga makna sitasi secara bersamaan. Gambar 4.1. menyajikan gambar arsitektur model terbaik.



Gambar 4.1. Arsitektur model terbaik MTL SP CNN

5. BOBOT ATRIBUT, KERANGKA PENILAIAN KUALITAS SITASI, DAN PROTOTIPE SISTEM APLIKASI

5.1. Bobot atribut sitasi

Metode BWM merupakan salah satu metode subyektif untuk menentukan bobot atribut. Sebagai hasil riset empiris, bobot yang dihasilkan dari metode BWM dapat dikategorikan sebagai kebenaran dasar (*ground truth*). Nilai kebenaran dasar merupakan nilai rujukan untuk tujuan perbandingan standar (Cardoso et al., 2014). MCDM secara obyektif memberikan bobot skor berdasarkan karakteristik data. Setiap metode memiliki berbagai teknik, berbeda algoritma dan kompleksitas, serta pertimbangan tertentu dalam menentukan nilai/bobot

kriteria (sitasi crispyn). Perbandingan bobot yang dihasilkan oleh beberapa metode obyektif dapat memperlihatkan metode yang tepat sesuai dengan karakteristik data yang digunakan. Penelitian ini membandingkan bobot yang dihasilkan lima metode obyektif dengan bobot *ground truth* hasil BWM. Perbandingan nilai korelasi bobot *ground truth* dengan hasil lima metode obyektif dapat disajikan pada Tabel 5.1.

Tabel 5.1. Korelasi Spearman antara bobot hasil MCDM obyektif dengan bobot BWM

Korelasi dengan BWM	Metode MCDM obyektif				
	STD	Koef Gini	Entropi	CRITIC	MEREC
Bobot lokasi	0,041	-0,350	-0,460	-0,120	0,035
Bobot sentimen	-0,079	0,870	-0,650	0,800	0,850
Bobot sumber-peran	-0,640	0,550	0,870	0,570	0,760
Bobot fungsi	-0,670	0,490	0,062	0,540	0,490
Bobot total	0,099	0,670	0,570	0,710	0,710

Hasil perbandingan bobot, korelasi, dan peta menunjukkan setiap metode memiliki pola bobot yang berbeda-beda. Berdasarkan nilai korelasi, hanya ada satu metode yang nilai korelasi pada bobot atribut dan bobot keseluruhan bernilai positif yaitu MEREC. Berdasarkan nilai bobot, atribut sentimen dan fungsi pada MEREC menempatkan urutan bobot tertinggi seperti BWM yaitu bobot sentimen tertinggi adalah positif dan bobot fungsi tinggi adalah pembanding. Pada dua atribut lainnya, yaitu lokasi dan sumber peran, terdapat perbedaan penempatan namun masih sesuai dengan konsep tingkat kepentingan atribut analisis sitasi dalam teks, yaitu bobot tertinggi lokasi adalah pada pembahasan dan bobot sumber peran tertinggi adalah data (Athar dan Teufel, 2012; Maricic et al., 1998). Berdasarkan pertimbangan tersebut diputuskan metode obyektif yang tepat untuk pembobotan sitasi dalam teks adalah MEREC.

5.2. Kerangka penilaian kualitas sitasi

Penelitian ini mengusulkan kerangka penilaian kualitas sitasi yang mengukur tingkat kepentingan artikel yang disitir dalam artikel yang menyitir. Nilai kualitas sitasi ditentukan oleh lima variabel yaitu intensitas kemunculan, bobot lokasi, bobot sentimen, bobot sumber peran, dan bobot fungsi (Gambar 5.1).

Terdapat tiga perhitungan nilai kualitas sitasi. Pertama adalah nilai kualitas per kemunculan (NKS_f) yang menjumlahkan empat bobot atribut yaitu bobot lokasi, bobot sentimen, bobot sumber peran, dan bobot fungsi dari satu frekuensi kemunculan. Kedua adalah nilai kualitas per dokumen artikel menyitir (NKS_d). Nilai ini merupakan penjumlahan dari NKS_f untuk semua kemunculan dalam satu dokumen artikel yang sama. Ketiga adalah nilai kualitas sitasi total dalam seluruh ketersediaan data (NKS_T). Nilai ini merupakan penjumlahan dari NKS_d dari seluruh dokumen. Ketiga rumus nilai kualitas sitasi dinyatakan dalam persamaan 6.1-6.3. Sejauh pengetahuan penulis, kerangka ini merupakan yang pertama memasukan variabel atribut lokasi dan tiga makna konteks sitasi untuk evaluasi kinerja riset berbasis sitasi. Metrik lain yang paling mendekati adalah metrik SJI oleh Nicholson et al. (2021) namun hanya menilai sitasi dari variabel sentimen.

- 1) Nilai kualitas sitasi untuk satu frekuensi kemunculan (NKS_f)

$$NKS_f = \sum L_f + S_f + P_f + F_f$$

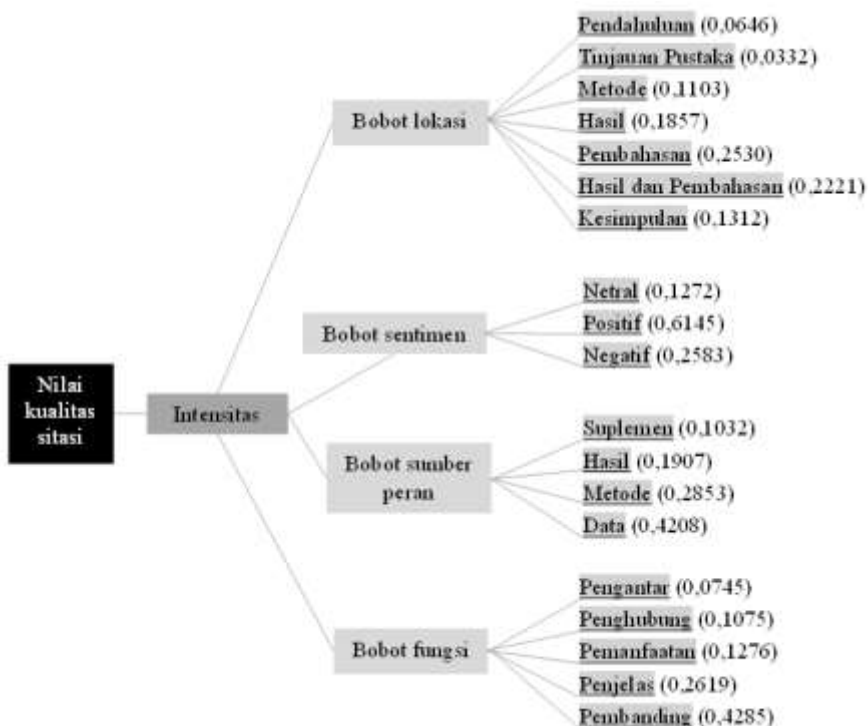
- 2) Nilai kualitas sitasi untuk satu dokumen artikel yang menyitir (NKS_d)

$$NKS_d = \sum_{i=1}^n (NKS_f)_i$$

- 3) Nilai kualitas sitasi total (NKS_T)

$$NKS_T = \sum_{j=1}^n (NKS_d)_j$$

L_f = bobot lokasi dari satu kemunculan
 S_f = bobot sentimen dari satu kemunculan
 P_f = bobot sumber peran dari satu kemunculan
 F_f = bobot fungsi dari satu kemunculan
 i = indeks sitasi dalam teks
 j = indeks dokumen menyitir
 n = frekuensi



Gambar 5.1. Kerangka penilaian kualitas sitasi

Salah satu variabel utama dalam penghitungan NKS adalah intensitas (frekuensi) kemunculan sitasi dalam teks. Variabel ini rawan penyalah-gunaan dimana penulis dapat

menuliskan sitasi beberapa kali untuk meningkatkan NKS padahal sebenarnya memiliki makna yang sama. Oleh karena itu, untuk mengurangi potensi kecurangan, dilakukan normalisasi pada variabel intensitas dalam NKS. Sitasi yang sama ketika muncul lebih dari satu kali di lokasi yang sama dan memiliki makna yang sama (sentimen, sumber peran, dan fungsi) maka dihitung sebagai satu kali kemunculan. Proses normalisasi membuat NKS lebih fokus mengukur kualitas sitasi berdasarkan maknanya dengan tetap memperhitungkan kemunculan yang bermakna.

Evaluasi terhadap metrik NKS dilakukan melalui perbandingan ranking sitasi yang dihasilkan oleh metrik NKS dengan ranking yang ditentukan oleh ahli. Hasil korelasi terhadap 10 artikel sampel disajikan pada Tabel 6.2. Dua ahli yang menilai sitasi dalam sepuluh dokumen sampel memiliki nilai korelasi yang mirip pada sembilan dokumen. Hanya pada satu dokumen, artikel ke-8, ada perbedaan cukup besar dimana ahli satu mendapatkan nilai korelasi 0,1839, sedangkan ahli dua memperoleh nilai korelasi 0,6317. Kemiripan ini mengindikasikan kedua ahli memiliki pendapat yang hampir sama dalam menilai sitasi. Adapun nilai rata-rata korelasi adalah 0,64. Nilai ini menunjukkan ada hubungan yang kuat antara penilaian NKS dengan penilaian ahli.

Tabel 5.2. Korelasi Spearman antara penilaian hasil NKS dan penilaian ahli

Nilai korelasi	Sampel artikel									
	1	2	3	4	5	6	7	8	9	10
Ahli 1	0,62	0,51	0,52	0,97	0,51	0,40	0,82	0,18	0,89	0,53
Ahli 2	0,53	0,49	0,76	1,00	0,52	0,52	0,93	0,63	1,00	0,48
Rata-rata nilai korelasi	0,64									

5.3. Prototipe sistem aplikasi

Gambaran mendalam mengenai pemanfaatan hasil klasifikasi makna sitasi, pembobotan atribut, dan kerangka penilaian kualitas sitasi ditampilkan dalam prototipe sistem

aplikasi berbasis situs web. Sistem terdiri atas tiga menu yaitu (1) informasi sitasi artikel, (2) informasi makna sitasi, dan (3) nilai kualitas sitasi.

Menu pertama, informasi sitasi artikel, berisi fitur pencarian artikel dan informasi sitasinya. Sebagai halaman beranda, pada *header* halaman ditampilkan informasi ringkas tentang analisis sitasi dalam teks dan atributnya. Pencarian artikel dapat dilakukan oleh pengguna dengan mengisi salah satu atau gabungan dari empat opsi inputan yaitu (i) penulis, (ii) tahun, (iii) judul, dan (iv) sumber artikel yang dicari. Keluaran dari pencarian ini adalah tabel berisi informasi bibliografi artikel (judul, tahun, penulis, sumber), nilai kualitas sitasi, dan tautan informasi detail makna sitasi dari artikel (Gambar 5.2).

Menu kedua adalah informasi makna sitasi. Menu ini untuk memfasilitasi pengguna yang ingin mencari artikel dari maknanya. Pengguna dapat mencari artikel berdasarkan lokasi sitasi, sentimen sitasi, sumber peran sitasi, dan atau makna sitasi. Bagian *header* halaman ini juga menampilkan ilustrasi atribut dan makna sitasi, sehingga pengguna mendapatkan gambaran input yang bisa dimasukkan. Input dapat dipilih satu-satu atau gabungan dua, tiga, empat input. Hasil dari pencarian ini adalah data artikel disitir, artikel menyitir, lokasi, dan tiga makna konteks kalimat sitasi (sentimen, sumber peran, dan fungsi)(Gambar 5.3).

Menu ketiga dalam prototipe adalah nilai kualitas sitasi. Pada menu ini, *header* halaman antarmuka menampilkan definisi dan rumus penilaian kualitas sitasi. Ada tiga nilai yang disajikan yaitu jumlah sitasi, jumlah sitasi dalam teks, dan nilai kualitas sitasi total. Pada bagian bawah tersaji nilai kualitas sitasi dari seluruh artikel berurutan dari nilai tertinggi (Gambar 5.4). Pengguna juga dapat mencari/memfilter dengan mengisi input bibliografi artikel, sehingga hanya akan menampilkan nilai kualitas sitasi dari artikel yang dicari. Informasi dari halaman ini dapat menjadi pertimbangan dalam rekomendasi sitasi dan penilaian artikel untuk evaluasi kinerja riset.

CARI DARI ARTIKEL		CARI DARI MAKNA		NILAI KUALITAS SITASI	
Penulis	Tahun	Judul	Sumber		
<input type="text"/>	2019	<input type="text"/>	<input type="text"/>		
<input type="button" value="SEARCH"/>					
Penulis	Tahun	Judul	Sumber	Nilai kualitas sitasi total (NKST)	Makna sitasi
Aini, Q., Rehsrdia, U., Arribathi, A. H., & Santoso, N. P. L.	2019	Penerapan Cloud Accounting dalam Menunjang Efektivitas Laporan Neraca pada Perguruan Tinggi	CESS (Journal of Computer Engineering, Science and System), 4(1), 60-64	0.4482	Detail makna sitasi
AMINUDIN, AMINUDIN	2019	Analisa Performa Apache Hadoop Dengan H2o Menggunakan Benchmark Hibench Via Cloud Computing	Prosiding SENTRA (Seminar Teknologi Dan Rekayasa)	0.3695	Detail makna sitasi
Febriyanto, E., Handayani, I., & Suprayogi, D.	2019	Aplikasi Sistem Penilaian Penguji Berbasis YII Framework Sebagai Media Input Nilai Mahasiswa Sidang Tugas Akhir dan Skripsi pada Perguruan Tinggi	CSRID (Computer Science Research and Its Development Journal), 10(2), 113-125	0.4482	Detail makna sitasi
Handayani, I., Yulianto, & Khumaida, FA.	2019	Sistem Penilaian Penguji Pada PESSITA+S2 Berbasis YII Framework Sebagai Media Penginputan Nilai Sidang Tesis di Perguruan Tinggi	SENSI Journal (Strategic of Education in Information System Journal), 5(1), 63-72	0.4152	Detail makna sitasi
Handriani, A. N. S., Sidik, A. Mulfazam, and I. Wijaya	2019	-	Ensiklopedia of Journal, vol. 1, no. 2, pp. 117-125	0.3695	Detail makna sitasi
Kurniawan, A.	2019	Definisi, Pengertian Teknologi Menurut Para Ahli Dan Perkembangannya.	-	0.3695	Detail makna sitasi

Gambar 5.2. Contoh hasil pencarian dari tahun terbit artikel

CARI DARI ARTIKEL CARI DARI MAKNA NILAI KUALITAS SITASI

Cari berdasarkan atribut lokasi dan makna, untuk memperoleh rekomendasi sitasi yang diinginkan :

Lokasi: Sentimen: Sumber peran: Fungsi:

SEARCH

Artikel yang disitir	Artikel yang menyitir	Kalimat	Lokasi	Sentimen	Sumber peran	Fungsi	Nilai kualitas sitasi per kemunculan (NKSI)
Lowy, "Object Recognition from Local Scale-Invariant Features," in Proceedings of the Seventh IEEE International Conference on Computer Vision, Kerkyra, Greece, Sept. 1999, pp. 1150-1152.	Hadha Afrisal. 2019. Metode Pengenalan Tempat Secara Visual Berbasis Fitur CNN untuk Navigasi Robot di Dalam Gedung. Jurnal Teknologi dan Sistem Komputer, 7(2): 47-55	Pada eksperimen yang dilakukan, metode SIFT yang digunakan yang berbeda dengan metode SIFT original yang dikembangkan []	Metode	Negatif	Metode	Pembandingan	1.0824
Omase, E. N. Ndungu, P. L. Kibet, and H. Tarus, "Artificial Intelligence approach to signal propagation modeling for outdoor to indoor wireless communication networks; a proposed study," in the Mechanical Engineering Conference on Sustainable Research and Innovation, Nairobi, Kenya, May 2012, pp. 289-299.	Fransiska Sisilia Mukti. 2019. Studi komparatif empat model propagasi empiris dalam ruangan untuk jaringan nirkabel kampus. Jurnal Teknologi dan Sistem Komputer, 7(4): 154-160	. Model ini tidak memperhitungkan keberadaan obstacle, melainkan hanya faktor power decay saja yang dipertimbangkan []	Metode	Negatif	Metode	Penghubung	0.7614
FENNY AND ANDRY, J. F. (2017) 'Audit sistem informasi menggunakan framework cobit 4.1 pada pt. anisa solusi teknologi', (November).	Innoatul Muthoharoh, Effara Fraksi Dwijpa, Renny Sari Dewi. 2019. Analisis Tingkat Kematangan Pendistribusian Ruang Kelas Dengan Framework Cobit 4.1 Menggunakan Domain Ds10, Jurnal Teknologi Informasi dan	Tabel 1. Penelitian terdahulu COBIT 4.2. Tidak difokuskan pada satu domain. Celah: Tidak memiliki laporan temuan dan rekomendasi untuk studi kasus tersebut []	Metode	Negatif	Metode	Penghubung	0.7614

Gambar 5.3. Hasil pencarian makna sitasi

	CARI DARI ARTIKEL	CARI DARI MAKNA	NILAI KUALITAS SITASI
Artikel yang diteliti			Jumlah sitasi Sitasi dalam teks Nilai kualitas sitasi total (NQI) ▼
Oliva, E. N. Ndungu, R. L. Kibet, and H. Tiers, "Artificial intelligence approach to signal propagation modeling for outdoor to indoor wireless communication networks: a proposed study," in the Mechanical Engineering Conference on Sustainable Research and Innovation, Nairobi, Kenya, May 2017, pp. 289-299.			1 8 4,209
Meyer and J. P. Harrison, "Impact of Single Parameter Changes on Ceph Cloud Storage Performance," <i>Scalable Computing</i> , vol. 17, no. 4, pp. 285-298, 2013.			1 3 4,19
Sunarya, E. Y. Mujiyanto, M. H. Partomo, T. A. Saripono, I. Sunu, and I. K. E. Purnama, "Carotid Artery B-Mode Ultrasound Image Segmentation based on Morphology, Geometry and Gradient Direction," in Second International Workshop on Pattern Recognition, Singapore, Jun. 2017, pp. 3-9.			1 9 4,0011
Yoon, K. H. Nam, C. Ju, D. G. Pwng, and S. J. Lee, "3D Reconstruction of a Carotid Bifurcation from 2D Transverse Ultrasound Images," <i>Ultrasonics</i> , vol. 54, no. 11, pp. 2104-2107, 2014.			1 6 4,0611
Vegas, H. Martha Fabiola Cortezas, and B. Jose Antonio Vegas, "Design of empirical propagation models supported in the log-normal shadowing model for the 2.4GHz and 5GHz bands under indoor environments," <i>Indian Journal of Science & Technology</i> , vol. 11, no. 22, pp. 1-18, 2018. doi: 10.17485/ijst/2018/v11i22/122149			1 11 3,9374
Johari, H. F. Ihsaid, M. Nizam, M. Mylis, and N. Wilco, "Comparison of Various Virtual Machine Disk Images Performance on GlusterFS and Ceph Redox Block Devices," in the 3rd International Conference on Informatics & Applications, Kuala Terengganu, Malaysia, 2014, pp. 1-7.			1 4 3,8425
Promchan, N. Napsakul, and K. Promsopa, "Performance Evaluation of ZFS and LVM (with ext4) for Scalable Storage System," in 2011 8th International Joint Conference on Computer Science and Software Engineering, Nakhon Pathom, Thailand, May 2011, pp. 250-253.			1 4 3,6604
Pleasantang, H. Manerjilang, and E. Zylmusani, "Optimizing Support Vector Machine Parameters with Genetic Algorithm for Credit Risk Assessment," <i>Journal of Physics: Conference Series</i> , vol. 930, no. 1, 2017.			1 4 3,5239
Muarqob, "Komparasi Algoritma Klasifikasi Data Mining untuk Evaluasi Pemberian Kredit," <i>Bina Informatika Journal</i> , vol. 3, no. 1, pp. 187-193, 2016.			1 4 3,5236
Liu, "Indoor radio propagation modeling for system performance prediction," Thesis, INSA de Lyon, France, 2013.			1 11 3,3308
Chen and T. Koz, "Performance evaluation of IoT protocols under a constrained wireless access network," in 2016 International Conference on Selected Topics in Mobile & Wireless Networking (MoWNet), Cairo, Egypt, Apr. 2016, pp. 1-7. doi:10.1109/MoWNet.2016.7496022			1 3 3,14
Yıldız, I. Kubela, S. Basar, and K. Tatavsalı, "Performance evaluation of application layer protocols for the internet of things," in 2018 Sixth International Conference on Enterprise Systems (ES), Limassol, Cyprus, Oct. 2018, pp. 180-187. doi:10.1109/ES.2018.00035			1 3 3,14

Gambar 5.4. Nilai kualitas sitasi seluruh artikel

6. PENUTUP

6.1. Kesimpulan

Penelitian ini bertujuan untuk mengembangkan model MTL untuk mengklasifikasi tiga makna sitasi secara bersamaan. Eksperimen klasifikasi menggunakan enam belas model terdiri atas sembilan model MTL, tiga model *deep learning* tugas tunggal, dan empat model mesin pembelajar tradisional. Hasil evaluasi kinerja menemukan model terbaik adalah model MTL *shared-private layer* CNN. Model terbaik ini memiliki skor $F1 > 0,80$ untuk tiga tugas klasifikasi tiga makna sitasi dengan skor tertinggi untuk sumber peran (0,87) dan fungsi (0,90), namun untuk klasifikasi sentimen memperoleh skor (0,81) lebih rendah dari CNN tugas tunggal (0,85). Karakteristik data yang tidak seimbang menjadi tantangan utama khususnya pada klasifikasi sentimen dan ketepatan deteksi kelas minoritas.

Tujuan penelitian kedua adalah menentukan pembobotan atribut sitasi dan kerangka pengukuran nilai kualitas sitasi dalam artikel jurnal. Konsep kualitas sitasi dalam teks memberikan bobot berbeda pada lokasi, sentimen, sumber peran, dan fungsi. Penentuan bobot kebenaran dasar menggunakan metode subyektif MCDM *best worst method*. Perbandingan bobot *ground truth* dengan bobot lima metode obyektif MCDM menetapkan bobot hasil metode MEREC paling mendekati dan memiliki nilai korelasi ranking paling tinggi. Kerangka penilaian kualitas sitasi mengusulkan metrik nilai kualitas sitasi (NKS) yang merupakan penjumlahan bobot lokasi, sentimen, sumber peran, dan fungsi dari seluruh kemunculan sitasi dalam teks. Prototipe sistem aplikasi berisi fitur pencarian artikel, makna sitasi, dan nilai kualitas sitasi dikembangkan untuk memberikan gambaran pemanfaatan hasil penelitian.

6.2. Keterbatasan penelitian

Penelitian ini membatasi cakupan pada artikel jurnal berbahasa Indonesia terakreditasi Sinta 1 dan Sinta. Saat ini banyak jurnal berkualitas di Indonesia terindeks internasional ditulis menggunakan bahasa Inggris. Adapun jurnal terindeks Sinta 1 dan 2 yang masuk dalam set data masih banyak yang tidak memenuhi standar penulisan yang baik. Kondisi dan temuan ini menyebabkan hasil penelitian tidak sepenuhnya menggambarkan pola sitasi dalam jurnal ilmiah di Indonesia.

Penelitian berfokus pada proses klasifikasi, pembobotan atribut sitasi dan penilaian kualitas sitasi. Beberapa proses dalam penyusunan set data menggunakan perangkat yang sudah tersedia. Perangkat yang ada dibuat untuk mengolah artikel jurnal internasional dan tidak sepenuhnya sesuai dengan karakteristik data jurnal di Indonesia. Hal ini menyebabkan tingkat akurasi yang tidak maksimal sehingga memerlukan proses pengolahan secara manual. Proses manual dilakukan pada hasil parsing lokasi dan konteks sitasi.

Eksperimen klasifikasi menggunakan set data pribadi karena belum ada set data publik yang memiliki label tiga makna sitasi untuk kalimat yang sama. Kondisi ini menyebabkan tidak bisa dilakukan perbandingan evaluasi model dengan data yang sudah digunakan dalam penelitian sebelumnya.

6.3. Saran penelitian selanjutnya

Keterbatasan penelitian dan upaya peningkatan kualitas hasil mendorong diperlukannya beberapa penelitian lanjutan. Perbaikan kuantitas dan kualitas set data dapat dilakukan dengan menambah cakupan pada jurnal penerbit Indonesia berbahasa Inggris. Proses manual pada penyusunan data memerlukan penelitian terkait alat parsing khusus artikel jurnal Indonesia dan metode penentuan doi/identitas unik secara otomatis. Penelitian pembangunan model klasifikasi perlu dilanjutkan dengan berfokus pada kombinasi data kalimat

bahasa Indonesia dan bahasa Inggris, mengatasi kelas data tidak seimbang, menggunakan metode terkini dalam model klasifikasi seperti model berbasis transformer, penggunaan fitur kualitas teks, dan evaluasi menggunakan metode yang lebih sensitif. Metrik nilai kualitas sitasi, berbasis atribut lokasi dan makna sitasi dalam teks, yang diusulkan dalam penelitian ini baru menghitung nilai sitasi pada level dokumen. Padahal penilaian kinerja umumnya dilakukan terhadap penulis, jurnal, atau lembaga. Oleh karena itu perlu dilakukan penelitian lanjutan untuk dapat menghitung kualitas sitasi dari penulis, jurnal, dan lembaga. Analisis dan klasifikasi makna relevansi/kemiripan antara kalimat sitasi dengan artikel yang disitir dapat ditambahkan dalam metode MTL dan variabel pembobotan dalam penelitian lebih lanjut untuk meningkatkan ketepatan penilaian kualitas sitasi. Implementasi dan pemanfaatan penelitian perlu dikembangkan dalam bentuk sistem aplikasi. Penelitian terkait pembangunan sistem, mulai dari penentuan kebutuhan pengguna, antar muka, dan evaluasi lainnya perlu dilakukan untuk melengkapi gambaran besar penelitian ini.

DAFTAR PUSTAKA

- Athar, A., & Teufel, S. (2012). Context-enhanced citation sentiment detection. *NAACL HLT 2012 - 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Proceedings of the Conference*, 597–601.
- Cardoso, J. R., Pereira, L. M., Iversen, M. D., & Ramos, A. L. (2014). What is gold standard and what is ground truth? *Dental Press Journal of Orthodontics*, 19(5), 27–30. <https://doi.org/10.1590/2176-9451.19.5.027-030.ebo>
- Crawshaw, M. (2020). Multi-task learning with deep neural networks: A survey. *ArXiv Preprint ArXiv: 2009.09796*.

- De Bellis, N. (2010). *Bibliometrics and Citation Analysis: From the Science Citation Index to Cybermetrics* (Vol. 23, Issue 3). The Scarecrow Press, Inc. <https://doi.org/10.1087/20100312>
- Gingras, Y. (2014). *Bibliometrics and research evaluation*. MIT Press.
- Lever, J., Krzywinski, M., & Altman, N. (2016). Classification evaluation. *Nature Methods*, 13(8), 603–605.
- Liu, P., Qiu, X., & Huang, X. (2017). Adversarial multi-task learning for text classification. *ArXiv*, 1704.05742.
- Lopez, P. (2009). GROBID: Combining automatic bibliographic data recognition and term extraction for scholarship publications. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 5714 LNCS, 473–474. https://doi.org/10.1007/978-3-642-04346-8_62
- Maricic, S., Spaventi, J., Pavicic, L., & Pifat-mrzljak, G. (1998). Citation Context versus the Frequency Counts of Citation History. *Journal of American Society for Information Science*, 49(6), 530–540. [https://doi.org/https://doi.org/10.1002/\(SICI\)1097-4571\(19980501\)49:6<530::AID-ASI5>3.0.CO;2-8](https://doi.org/https://doi.org/10.1002/(SICI)1097-4571(19980501)49:6<530::AID-ASI5>3.0.CO;2-8)
- Misra, I., Shrivastava, A., Gupta, A., & Hebert, M. (2016). Cross-stitch networks for multi-task learning. *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- Ruder, S. (2017). An Overview of Multi-Task Learning in Deep Neural Networks. *ArXiv Preprint ArXiv:1706.05098*.
- Shahid, A., Afzal, M. T., & Qadir, M. A. (2015). Lessons learned: The complexity of accurate identification of in-text citations. *International Arab Journal of Information Technology*, 12(5), 481–488.

- White, H. D. (2004). Citation Analysis and Discourse Analysis Revisited. *Applied Linguistics*, 25(1), 89–116. <https://doi.org/10.1093/applin/25.1.89>
- Yaniasih, Y. (2020). Teori kritis terhadap analisis sitasi untuk kajian kuantitatif sains dan evaluasi kinerja riset. *Berkala Ilmu Perpustakaan Dan Informasi*, 16(1), 127–141. <https://doi.org/10.22146/bip.v16i1.72>
- Zhang, Y., & Yang, Q. (2021). A Survey on Multi-Task Learning. *IEEE Transactions on Knowledge and Data Engineering*, 4347(c), 1–20. <https://doi.org/10.1109/TKDE.2021.3070203>

Publikasi:

1. Yaniasih, Y. (2020). Teori kritis terhadap analisis sitasi untuk kajian kuantitatif sains dan evaluasi kinerja riset. *Berkala Ilmu Perpustakaan Dan Informasi*, 16(1), 127–141.
2. Yaniasih, Y., & Budi, I. (2021). Analysis of In-text Citation Patterns in Local Journals for Ranking Scientific Documents. *DESIDOC Journal of Library & Information Technology*, 41(2), 94–101.
3. Yaniasih, Y., & Budi, I. (2021). Systematic Design and Evaluation of a Citation Function Classification Scheme in Indonesian Journals. *Publications*, 9(27), 1–14.
4. Budi, I., & Yaniasih, Y. (2022). Understanding the meanings of citations using sentiment, role, and citation function classifications. *Scientometrics*, 8.