

Klasifikasi Pekerjaan Bidang Teknologi Informasi Menggunakan Algoritma Cosine Similarity

Puji Catur Siswipraptini^{1*)}; Muhammad Istigfahri Wafit¹; Karina Djunaidi¹

1. Fakultas Telematika Energi, Institut Teknologi PLN, DKI Jakarta 11750, Indonesia

^{*)}Email: pujicatur@itpln.ac.id

Received: 24 April 2023 / Accepted: 24 April 2023 / Published: 27 April 2023

ABSTRACT

Skills for each type of work must be prepared since students are in higher education. The type of work and its supporting skills become one of the foundations for the study program in formulating the specialization of courses according to the needs of the industry. So the study program needs guidelines to develop a curriculum that is in accordance with industry needs. This study aims to classify the types of the Information Technology (IT) jobs field according to the 2020 Computing Curricula issued by ACM as one of the guidelines for Higher Education in the field of Computer Science in Indonesia in compiling graduate competence and curriculum. Job classification is carried out based on the needs of the industrial world as reflected in job vacancy advertisements and websites in Indonesia. The classification technique is carried out in several steps, namely web scraping on job vacancy websites; preprocessed text to clean web scraping data; and classification of 10 types of jobs using the Cosine Similarity algorithm. The cosine similarity algorithm analyzes the distance between documents for all types of work with required skills for job classification. The highest similarity level is 63% in Computer Network Architects class and the highest accuracy value of 33,5 % using k-fold cross validation.

Keywords: Information Technology jobs, Cosine Similarity, k-Fold Cross Validation

ABSTRAK

Keterampilan setiap jenis pekerjaan harus disiapkan sejak mahasiswa berada di jenjang pendidikan tinggi. Jenis pekerjaan dan keterampilan pendukungnya menjadi salah satu landasan bagi program studi dalam menyusun peminatan mata kuliah sesuai kebutuhan industri. Sehingga program studi membutuhkan pedoman untuk menyusun kurikulum yang sesuai dengan kebutuhan industri. Penelitian ini bertujuan melakukan klasifikasi jenis pekerjaan di bidang Teknologi Informasi (TI) menurut Computing Curricula tahun 2020, yang dikeluarkan oleh ACM sebagai salah satu pedoman Perguruan Tinggi (PT) di bidang Ilmu Komputer yang ada di Indonesia dalam menyusun kompetensi lulusan dan kurikulum PT. Klasifikasi pekerjaan dilakukan berdasarkan kebutuhan dunia industri yang tercermin pada iklan/website lowongan pekerjaan yang ada di Indonesia. Teknik klasifikasi dilakukan dengan beberapa langkah yaitu web scraping pada website lowongan pekerjaan, teks pra prosesing untuk membersihkan data hasil web scraping dan klasifikasi 10 jenis pekerjaan menggunakan algoritma Cosine Similarity. Algoritma Cosine Similarity melakukan analisis jarak kemiripan antar dokumen keseluruhan jenis pekerjaan dengan keterampilan yang dibutuhkan pada setiap pekerjaan bidang Teknologi Informasi. Tingkat kemiripan tertinggi sebesar 63% pada pekerjaan Computer Network Architects dan akurasi tertinggi diperoleh sebesar 33,5% menggunakan teknik k-fold cross validation.

Kata kunci: Pekerjaan bidang Teknologi Informasi, Cosine Similarity, k-Fold Cross Validation

1. PENDAHULUAN

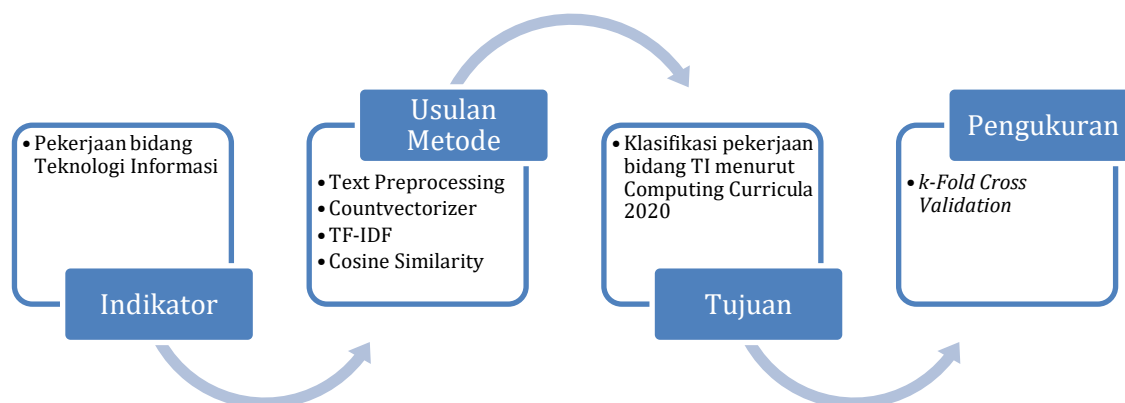
Pekerjaan bidang Teknologi Informasi (TI) sangat banyak di temui pada berbagai bidang industri. Asosiasi Pendidikan Tinggi Informatika dan Komputer (APTIKOM) di Indonesia telah merilis ada sekitar 550 pekerjaan bidang TI yang telah diidentifikasi [1]. Era teknologi industri 4.0 dan 5.0 membuat kebutuhan sumber daya manusia yang memiliki kompetensi di bidang TI semakin meningkat di berbagai industri maupun hampir seluruh kehidupan manusia [2]. Beragamnya jenis pekerjaan bidang TI memerlukan klasifikasi jenis pekerjaan yang sesuai dengan kompetensi yang dibutuhkan. Penelitian ini bertujuan untuk melakukan klasifikasi jenis pekerjaan TI menggunakan algoritma *Cosine Similarity* berdasarkan *Computing Curricula* tahun 2020 sebagai landasan bagi Pendidikan Tinggi Informatika dan Komputer yang ada di Indonesia[3]. Integrasi teknik *text mining* digunakan pada penelitian ini karena teknik ini menjadi salah satu tren penelitian di bidang rekomendasi pekerjaan[4]. *Text mining* biasanya digunakan untuk pencarian informasi berupa teks yang diperoleh dari jejaring sosial seperti media sosial atau iklan. Peran jejaring sosial dalam kehidupan sehari-hari saat ini sangat berguna untuk banyak hal karena jumlah informasi yang tersedia dapat mencapai ribuan data, termasuk mencari pekerjaan yang tepat [5]–[7].

Algoritma *Cosine Similarity* digunakan untuk menghitung tingkat kesamaan dari dua objek dan merupakan takaran sudut antara vektor. Input berupa bobot dari *term* tiap data yang digunakan pada proses menghitung kesenjangan antara kemiripan dengan kata tiap klaster, selanjutnya tiap nilai akan menetapkan centroid di setiap klaster[8]. Implementasi *Cosine Similarity* terbukti mampu memiliki akurasi yang sangat baik untuk menghitung kesamaan antara dua dokumen berdasarkan sudut Cosinus[9] di berbagai bidang seperti bidang pendidikan untuk mendeteksi kemiripan judul skripsi[10], rekomendasi seminar[11], sistem pendukung keputusan dengan multi kriteria[12] dan klasifikasi sentimen[13].

2. METODE PENELITIAN

2.1. Kerangka Pemikiran

Kerangka pemikiran pada penelitian ini seperti terlihat pada Gambar 1.



Gambar 1. Kerangka Pemikiran Klasifikasi Pekerjaan bidang TI

Berdasarkan Gambar 1 dapat dijelaskan sebagai berikut:

1. Indikator

Indikator sebagai input (masukan) berupa pekerjaan pekerjaan bidang TI yang di peroleh melalui proses Web Scraping pada website pekerjaan TechinAsia.

2. Usulan Metode

Usulan metode merupakan penerapan langkah langkah yang dilakukan untuk mencapai tujuan penelitian. Usulan metode terdiri atas beberapa langkah yaitu langkah pertama adalah *text preprocessing* untuk melakukan pembersihan dan normalisasi pada dataset yang diperoleh melalui proses web scraping. Langkah kedua adalah *countvectorizer* yaitu penghitungan kemunculan kata

pada dokumen. Langkah ketiga yaitu pembobotan tiap kata dengan menggunakan metode *Term Frequency-Inverse Document Frequency* (TF-IDF) yang umum digunakan pada penelitian Text Mining[14]. Langkah terakhir adalah penerapan algoritma *Cosine Similarity* untuk klasifikasi teks/dokumen dengan menghitung kemiripan antar dokumen menggunakan normalisasi panjang vektor.

3. Tujuan

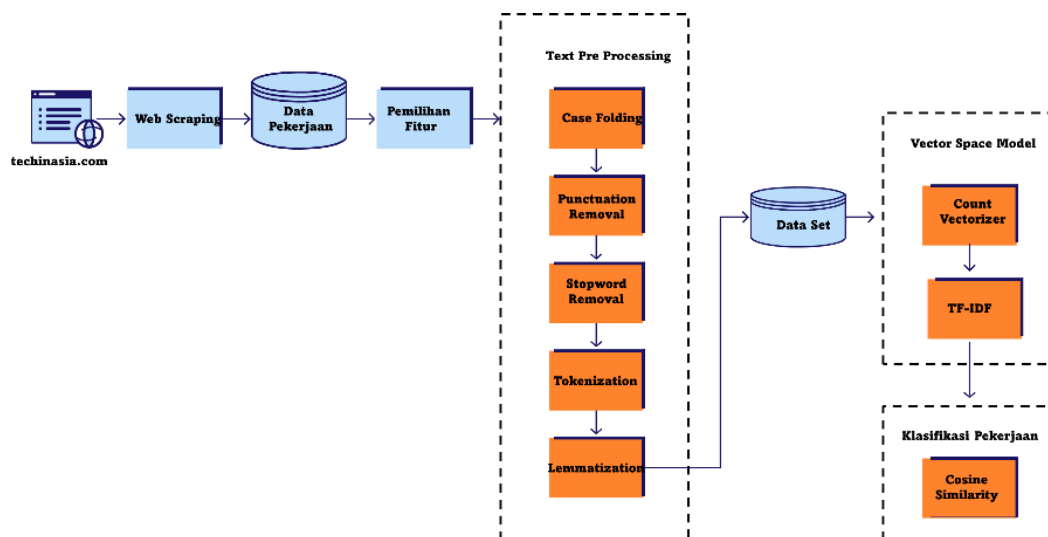
Tujuan akhir yang ingin dicapai pada penelitian ini adalah klasifikasi jenis pekerjaan bidang TI menjadi 10 kelas sesuai landasan pada dokumen *Computing Curricula 2020* [3] yang digunakan oleh pendidikan tinggi informatika dan ilmu komputer di Indonesia.

4. Pengukuran

Pengukuran dilakukan untuk menguji hasil penelitian yang telah dilakukan dengan menggunakan teknik *Cross Validation*.

2.2. Desain Penelitian

Desain penelitian seperti terlihat pada diagram alur gambar 2. Desain penelitian menjelaskan tahapan penggunaan beberapa metode untuk melakukan klasifikasi pekerjaan bidang TI. *Web scraping* adalah proses pengumpulan dataset dari website yaitu TechinAsia yang berfokus pada iklan pekerjaan TI di Indonesia. Hasil dari web scraping yaitu data pekerjaan dan dipilih fitur jenis pekerjaan dan keterampilan yang dibutuhkan pada setiap pekerjaan untuk dilakukan tahap *text preprocessing*. *Text preprocessing* terdiri atas *case folding* (merubah huruf besar menjadi huruf kecil), *punctuation removal* (menghapus tanda baca dan special karakter lainnya), *stopword removal* (menghilangkan kata penghubung), *tokenization* (merubah kalimat menjadi token) dan *lemmatization* (merubah bentuk kata menjadi kata dasar). Setelah dataset dilakukan normalisasi, maka tahap berikutnya adalah membentuk *Vector Space Model* (VSM) untuk mengukur kemiripan dokumen [15] dengan menggunakan metode *CountVectorizer* dan *TF-IDF*. Tahap akhir untuk klasifikasi pekerjaan bidang TI adalah implementasi algoritma *Cosine Similarity*.



Gambar 2. Desain Penelitian Klasifikasi Pekerjaan bidang TI

3. HASIL DAN PEMBAHASAN

3.1. Web Scraping

Hasil pengambilan data menggunakan teknik *web scraping* dilakukan pada bulan Pebruari - Maret 2022 dan didapat 970 dokumen pada website TechinAsia. Fitur yang didapat adalah lokasi pekerjaan, gaji, deskripsi pekerjaan, keterampilan yang dibutuhkan dan jenis pekerjaan. Proses pengambilan data ini dilakukan dengan dua cara yaitu scraping seluruh pekerjaan bidang TI dan scraping tiap kelas pekerjaan menggunakan kata kunci pada masing masing kelas pekerjaan. Berdasarkan *Computing Curricula* tahun 2020 terdapat 10 kelas/kategori pekerjaan bidang TI yaitu

Web Developers, Computer System Analysts, Computer Programmers, Information Security Analysts, Software Developers-Quality Assurance Analyst, and Testers, Computer Support Specialists, Database Administrators and Architects, Computer and Information Research Scientists, Computer Network Architects, dan Network and Computer System Administrators. Contoh hasil web scraping dapat dilihat pada seperti Tabel 1.

Tabel 1. Contoh hasil *web scraping* pekerjaan bidang TI di TechinAsia

Nama Perusahaan	Lokasi	Gaji	Deskripsi Pekerjaan	Keterampilan yang dibutuhkan	Jenis Pekerjaan
AiChat	Jakarta, Indonesia	IDR 10,000,000 – 15,000,000	Work on a front-end development and enhancement of UX for a Enterprise chatbot platform. Requirements - Minimum 3 years experience in JavaScript (and HTML/CSS). - Experienced with ReactJS, Bootstrap3, jQuery, AngularJS - Experienced in automated testing. - Passionate working for a fast paced start-up experience. - "Get it done" attitude and background of delivering great work. - Excellent verbal and written communication skills. - Bonus: strong UX understanding & product design Benefits - Flexible Working Hours - Attractive Remuneration - Fun & Dynamic Working Culture	HTML5 & CSS3 React.js Javascript	Frontend Developer
SIRCLO	Banten, Indonesia	IDR 5,000,000 – 10,000,000	RESPONSIBILITIES: Design and craft assets for the company's design needs.	Graphic Design Adobe Creative Suite Illustrator Adobe Photoshop	Graphic Designer

			<p>Translate briefs into visually developed designs that solves the objectives.</p> <p>Working together with users in crafting design for certain projects, such as banners, social media, campaign, event, and other design-related needs.</p> <p>REQUIREMENTS: Bachelor's degree in Visual Communication Design.</p> <p>At least 1-3 years of experience as a Graphic Designer or other similar roles.</p> <p>Experienced in operating Adobe Creative Cloud, including Illustrator, Photoshop, and InDesign.</p> <p>Able to work individually and as part of a team.</p> <p>Prior experience in handling social media is a plus.</p>	Adobe Indesign Social Media	
PT Rekeningku Dotcom Indonesia	Jakarta, Indonesia	IDR 7,000,000 – 20,000,000	<p>Rekeningku.com is marketplace cryptocurrency. we are building trading application for Indonesian crypto enthusiast.</p> <p>We are looking for really talented candidate to building fast, secure, scalable application.</p> <p>We open from any engineering background as long as highly motivated and talented.</p> <p>Requirement: Have strong technical</p>	Backend Development Javascript PHP CryptoCurrency Golang	Backend Developer

			programming Golang / Java / Node JS, MYSQL, Kafka, Caching mechanism		
Indocyber Global Teknologi	Jakarta, Indonesia	IDR 9,000,000 – 22,000,000	About the job The ideal candidate will be familiar with the full software design life cycle. They should have experience in designing, coding, testing and consistently managing applications They should be comfortable coding in a number of languages and have an ability to test code in order to maintain high-quality code. Responsibilities: Design, code, test and manage various applications. Collaborate with engineering team and product team to establish best products. Follow outlined standards of quality related to code and systems. Develop automated tests and conduct performance tuning.	ASP.NET Microsoft SQL Server HTML5 & CSS3 SQL Server jQuery REST APIs C# Microsoft Visual Studio	.NET Developer

3.2. Text Preprocessing

Text preprocessing dilakukan dengan menggunakan platform *GoogleColab* berbasis bahasa pemrograman Python dengan melakukan import *Natural Language Toolkit* (NLTK) untuk pemodelan teks. Dari fitur fitur pada tabel 1 dipilih fitur jenis pekerjaan dan keterampilan yang dibutuhkan untuk mencapai tujuan penelitian dan dilakukan tahap *text preprocessing*. Contoh hasil dari *text preprocessing* dipresentasikan pada Tabel 2 berikut.

Tabel 2. Contoh Hasil *Text Preprocessing*

Jenis Pekerjaan	Hasil Text Preprocessing
Frontend Developer	html css react j javascript
Site Reliability Engineer (DevOps)	devops jenkins git

Research Analyst	business development partnership market research analysis business analysis communication skill english language project management
Graphic Designer	graphic design adobe creative suite illustrator adobe photoshop adobe indesign social medium
Mobile Lead Engineer	io design ios development apache android

3.3. CountVectorizer

CountVectorizer dilakukan untuk penghitungan kemunculan kata/term pada dokumen hasil web scraping pekerjaan pada 10 kelas untuk mendapatkan frekuensi kata yang paling sering muncul. Kata/term yang paling sering muncul diasumsikan sebagai kata kunci pada sebuah pekerjaan yang menunjukkan tingkat kepentingan kata tersebut pada kumpulan dokumen pada masing-masing kelas. Adapun untuk hasil *countvectorizer* dari semua kelas pekerjaan disajikan pada tabel 3.

Tabel 3. Hasil *CountVectorizer* 10 kelas pekerjaan TI

Jenis Pekerjaan	Kata yang paling sering muncul
<i>Database Administrator</i>	administration data databases linux management mysql oracle postgresql server sql
<i>Computer And Information Reserch Scientist</i>	analysis analytics data mining databases machine learning
<i>Computer Network Architects</i>	cisco cloud cyber engineering linux network networking security web
<i>Software Developer System Software</i>	apis backend css design development engineering frameworks golang html javascript jquery js mongodb mysql programming python react rest software sql
<i>Computer System Analyst</i>	adwords agile analysis architecture data development engineering linux mysql product reporting sql strategy warehouse
<i>Web Developer</i>	algorithms angularjs apis backend cloud computing css development framework git golang html java js json laravel mysql node nosql php python react rest ruby saas web
<i>Software Developer Applications</i>	agile application android apis design development ios java js mobile react
<i>Information Security Analysts</i>	analysis big database management security skills warehouse

Computer Support Specialists	analysis communication data development enterprise information management networking software solving support strategy troubleshooting technical
Network And Computer System Administrators	administration aws databases mongodb mysql network nosql postgresql redshift security server services software sql system vmware

3.4. Term Frequency – Inverse Document Frequency (TF-IDF)

Term Frequency Inverse Document Frequency (TF-IDF) berfungsi menggabungkan dua konsep untuk perhitungan bobot nilai frekuensi kemunculan suatu kata di dalam suatu dokumen tertentu. Frekuensi kemunculan kata (*term*) di dalam dokumen yang berarti menunjukkan seberapa penting kata itu di dalam dokumen tersebut. Formula matematika yang digunakan pada TF-IDF sebagai berikut :

$$IDF = \log \left(\frac{D}{DF} \right) \quad (1)$$

$$Weight = TF \times IDF \quad (2)$$

Dimana:

D = jumlah dokumen dalam dataset

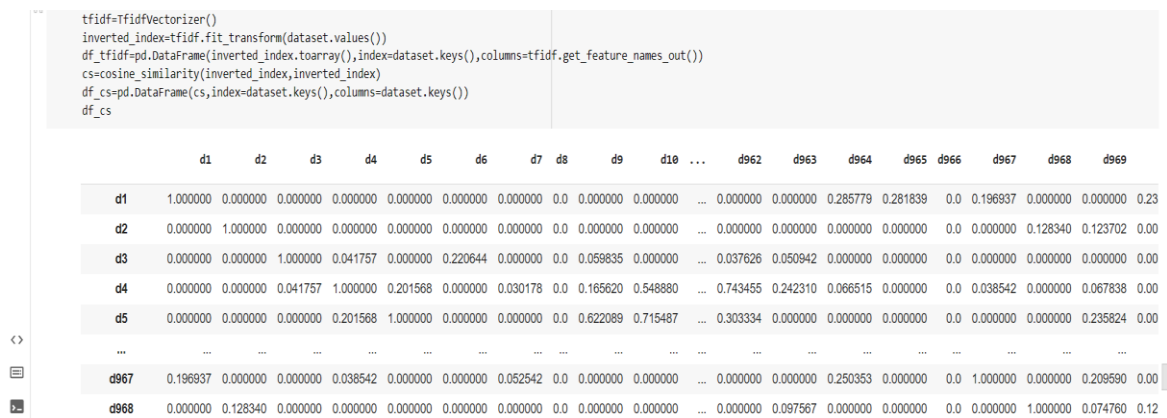
DF = jumlah dokumen yang mengandung kata (*term*)

W = hasil bobot nilai dokumen *Frequency*

TF = jumlah kata(*term*) dalam suatu dokumen

IDF = *Inverse Document Frequency*

Hasil dari TF-IDF dapat dilihat pada gambar 3.



Gambar 3. Hasil TF-IDF

3.4. Cosine Similarity

Cosine Similarity adalah salah satu algoritma di dalam teks mining, yang berfungsi untuk menghitung kesamaan antara dua buah objek yang dinyatakan dalam dua buah vector dengan menggunakan *keywords* (kata kunci). *Cosine similarity* mempunyai konsep normalisasi panjang vektor dengan membandingkan dua dokumen antara dokumen A dan B. Rumus yang umum digunakan untuk Cosine Similarity adalah :

$$similarity = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|} = \frac{\sum_{i=1}^n A_i B_i}{\sqrt{\sum_{i=1}^n A_i^2} \sqrt{\sum_{i=1}^n B_i^2}} \quad (3)$$

Dimana :

A = vektor A, yang akan dibandingkan kemiripannya

B = vektor B, yang akan dibandingkan kemiripannya

A.B = dot product antara vektor A dan vektor B

|A| = Panjang vektor A

|B| = Panjang vektor B

|A||B| = cross product antara vektor A dan vektor B

Contoh hasil dari Cosine Similarity pada pekerjaan Database Administrator dapat dilihat pada gambar 4.

```
[ ] #Pengurutan Ranking Cosine Similarity
rank_cs={}
for k in dataset.keys():
    if k != "Database Administrator":
        rank_cs[k]=df_cs.at[k,"Database Administrator"]
top_rank_cs=dict(sorted(rank_cs.items(),key=operator.itemgetter(1),reverse=True))
df_csr=pd.DataFrame(top_rank_cs.values(),index=top_rank_cs.keys(),columns=["Cosine Similarity"])
df_csr
```

	Cosine Similarity
d394	0.604122
d22	0.525103
d754	0.508356
d127	0.507867
d812	0.473217
...	...
d964	0.000000
d965	0.000000
d966	0.000000
d967	0.000000
d969	0.000000

Gambar 4. Contoh hasil Cosine Similarity pekerjaan Database Administrator

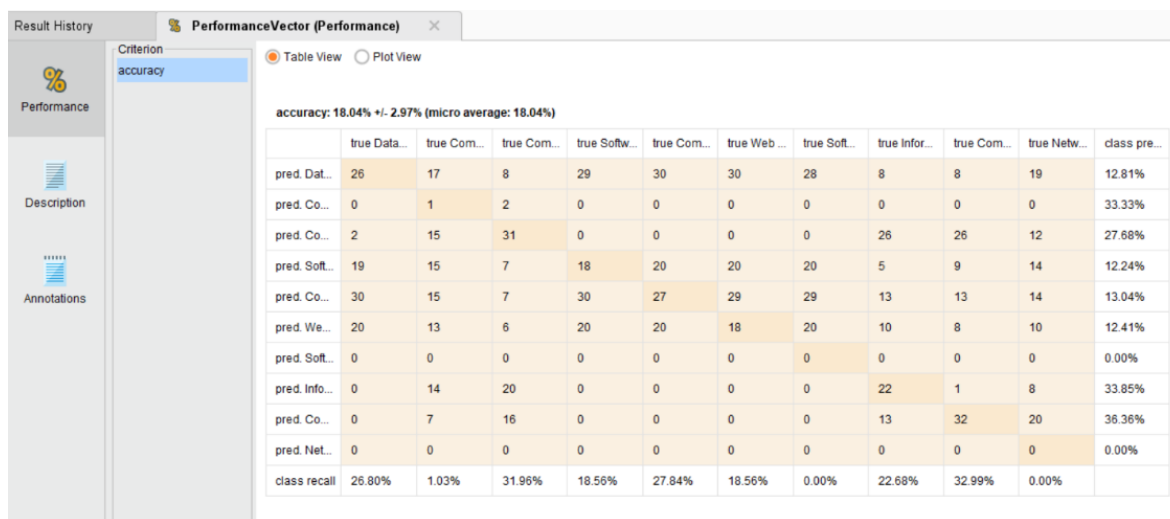
Dari hasil perhitungan menggunakan algoritma Cosine Similarity pada 10 pekerjaan TI di dapat nilai tertinggi kemiripan sebesar 63% pada pekerjaan *Computer Network Architects*. Seperti terlihat pada gambar 5 bahwa beberapa pekerjaan seperti *Senior Network Engineer*, *Digital Security*, *Senior Cyber Security Operation* dan lainnya di klasifikasi sebagai *Computer Network Architects*.

Nomor	Cosine Similarity	Jenis Pekerjaan	Required Skills
d661	0,63276049	Senior Network Engineer	cisco linux python network network s
d277	0,562596242	Digital Security	linux databases network security net
d259	0,48910079	Senior Cyber Security Operation	cyber security security information se
d829	0,486061013	IT Network Security	security cyber security network secur
d823	0,466643011	IT Security	cyber security information security n
d545	0,466084583	SYSTEM SECURITY ENGINEER	network security cyber security oper
d137	0,450384158	DevOps	linux amazon web service aws cloud
d112	0,4421875	IT Security Specialist	information security security network
d696	0,4421875	IT Security Specialist	information security security network
d190	0,430532544	Cyber Security Officer	cyber security infrastructure network
d781	0,388197638	Data Privacy (IT) Manager	information security security cyber se
d138	0,374816562	Network Engineer	network troubleshoot network secur
d421	0,310147028	Site Reliability Engineer	cloud compute linux git
d888	0,296179009	Odoo Developer	web development linux python
d708	0,28871672	FO Engineer	router network engineering
d292	0,285816447	Network Engineer (IT support)	information architecture network sec
d361	0,2819142	Network & System Engineer	infrastructure problem solve amazon
d422	0,278902727	IT Security Assistant Manager	devops linux cloud computing
d483	0,250390816	DevOps Engineer	automation javascript security cod cl
d204	0,24386392	DevOps Engineer	amazon web service aws linux system
d812	0,238181507	DevOps Engineer	devops cloud compute amazon web
d157	0,226560881	Tech Lead	java spring postgresql mysql nosql m
d75	0,22445956	Web Developer	html cs mysql javascript php web dev
d671	0,223321467	Senior DevOps Engineer	git amazon web service aws linux sys

Gambar 5. Klasifikasi pekerjaan *Computer Network Architects* menggunakan *Cosine Similarity*

3.5. *k-Fold Cross Validation*

k-Fold Cross Validation membagi dataset menjadi bagian terpisah dengan ukuran yang sama. Model dilatih oleh subset data latih dan divalidasi oleh subset validasi (data uji) sebanyak *k*. Pengujian model dilakukan menggunakan *K-Fold Cross Validation* menilai kinerja proses sebuah metode algoritma dengan membagi sampel data secara acak dan mengelompokkan data tersebut sebanyak nilai *K* *k-fold* dengan numbers of fold 10 dengan *sampling type automatic*. Kemudian salah satu kelompok 10-fold tersebut akan dijadikan sebagai data uji sedangkan sisa kelompok yang lain akan dijadikan sebagai data latih. Pengujian ini menggunakan *software Rapidminer Studio* dengan pembagian data training data testing sebesar 80:20. Hasil nya dapat dilihat pada gambar 6.



Result History | PerformanceVector (Performance) | Criterion: accuracy | Table View | Plot View

accuracy: 18.04% +/- 2.97% (micro average: 18.04%)

	true Data...	true Com...	true Com...	true Softw...	true Com...	true Web...	true Soft...	true Infor...	true Com...	true Netw...	class pre...
pred. Dat...	26	17	8	29	30	30	28	8	8	19	12.81%
pred. Co...	0	1	2	0	0	0	0	0	0	0	33.33%
pred. Co...	2	15	31	0	0	0	0	26	26	12	27.68%
pred. Soft...	19	15	7	18	20	20	20	5	9	14	12.24%
pred. Co...	30	15	7	30	27	29	29	13	13	14	13.04%
pred. We...	20	13	6	20	20	18	20	10	8	10	12.41%
pred. Soft...	0	0	0	0	0	0	0	0	0	0	0.00%
pred. Info...	0	14	20	0	0	0	0	22	1	8	33.85%
pred. Co...	0	7	16	0	0	0	0	13	32	20	36.36%
pred. Net...	0	0	0	0	0	0	0	0	0	0	0.00%
class recall	26.80%	1.03%	31.96%	18.56%	27.84%	18.56%	0.00%	22.68%	32.99%	0.00%	

Gambar 6. Hasil Akurasi, Rata-rata Presisi dan klasifikasi error pada model

Pada gambar 6 terlihat bahwa hasil uji validasi pada semua data yang telah di klasifikasi diperoleh hasil akurasi sebesar tertinggi sebesar 33,5 % dan presisi tertinggi sebesar 32,99 %.

4. KESIMPULAN DAN SARAN

Hasil penelitian membuktikan bahwa algoritma *cosine similarity* melakukan analisis jarak kemiripan antar dokumen 970 data *scraping* dengan 10 kategori pekerjaan bidang TI, tingkat kemiripan tertinggi sebesar 63% yaitu pada pekerjaan *Computer Network Architects*. Evaluasi model dengan metode *k-fold cross validation* mendapatkan hasil akurasi tertinggi sebesar 33,5%. Model yang diusulkan dapat digunakan oleh pencari kerja bidang Teknologi Informasi untuk mengetahui kompetensi yang harus dimiliki pada masing masing bidang pekerjaan tersebut.

DAFTAR PUSTAKA

- [1] APTIKOM, "Pengembangan Kurikulum KKNi berdasarkan OBE Bidang Ilmu Informatika dan Komputer," 2019.
- [2] Kementrian Tenaga Kerja, "Proyeksi Kebutuhan Tenaga Kerja Di Perusahaan Berdasarkan Kompetensi Pada Sektor Teknologi Informatika & Komunikasi Pada Tahun 2022 - 2025," 2021.
- [3] Association for Computing Machinery (ACM) and IEEE Computer Society (IEEE-CS), "Computing Curricula 2020 Paradigms for Global Computing Education," 2020.
- [4] P. C. Siswipraptini, H. L. H. Spits Warnars, A. Ramadhan, and W. Budiharto, "Trends and Characteristics of Career Recommendation Systems for Fresh Graduated Students," in *2022 10th International Conference on Information and Education Technology*, 2022, pp. 355–361.
- [5] I. A. Heggo and N. Abdelbaki, "Hybrid Information Filtering Engine for Personalized Job Recommender System," *Adv. Intell. Syst. Comput.*, vol. 723, pp. 553–563, 2018, doi: 10.1007/978-3-319-74690-6_54.

-
- [6] A. Rivas, P. Chamoso, A. González-Briones, R. Casado-Vara, and J. M. Corchado, "Hybrid job offer recommender system in a social network," *Expert Syst.*, vol. 36, no. 4, pp. 1–13, 2019, doi: 10.1111/exsy.12416.
- [7] G. Zhu, N. A. Kopalle, Y. Wang, X. Liu, K. Jona, and K. Börner, "Community-based data integration of course and job data in support of personalized career-education recommendations," in *Proceedings of the Association for Information Science and Technology*, 2020, vol. 57, no. 1, pp. 1–6, doi: 10.1002/pra2.324.
- [8] W. Firdaus, F., Pasnur, P., & Wabdillah, "Implementasi Cosine Similarity Untuk Peningkatan Akurasi Pengukuran Kesamaan Dokumen Pada Klasifikasi Dokumen Berita Dengan K-Nearest Neighbour," *J. Teknol. Inf. dan Komun.*, vol. 9, no. 1, pp. 69–74, 2019.
- [9] P. dwi Nurfadila, A. P. Wibawa, I. A. E. Zaeni, and A. Nafalski, "Journal Classification Using Cosine Similarity Method on Title and Abstract with Frequency-Based Stopword Removal," *Int. J. Artif. Intell. Res.*, vol. 3, no. 2, 2019, doi: 10.29099/ijair.v3i2.99.
- [10] F. E. P. Oppi Anda Resta, Addin Aditya, "View of Plagiarism Detection in Students' Theses Using The Cosine Similarity Method.pdf," *J. dan Penelit. Tek. Inform.*, vol. 5, no. 2, pp. 305–313, 2021.
- [11] M. Yusuf and A. Cherid, "Implementasi Algoritma Cosine Similarity Dan Metode TF-IDF Berbasis PHP Untuk Menghasilkan Rekomendasi Seminar," *J. Ilm. Fak. Ilmu Komput.*, vol. 9, no. 1, pp. 8–16, 2020, [Online]. Available: <https://publikasi.mercubuana.ac.id/index.php/fasikom/article/view/8830>.
- [12] E. Bolturk and C. Kahraman, "A novel interval-valued neutrosophic AHP with cosine similarity measure," *Soft Comput.*, vol. 22, no. 15, pp. 4941–4958, 2018, doi: 10.1007/s00500-018-3140-y.
- [13] T. Thongtan and T. Phienthrakul, "Sentiment classification using document embeddings trained with cosine similarity," *ACL 2019 - 57th Annu. Meet. Assoc. Comput. Linguist. Proc. Student Res. Work.*, pp. 407–414, 2019, doi: 10.18653/v1/p19-2057.
- [14] Y. Asri, W. N. Suliyanti, D. Kuswardani, and M. Fajri, "Pelabelan Otomatis Lexicon Vader dan Klasifikasi Naive Bayes dalam menganalisis sentimen data ulasan PLN Mobile," *PETIR J. Pengkaj. dan Penerapan Tek. Inform.*, vol. 15, no. 2, pp. 264–275, 2022.
- [15] R. Arianto, S. Warnars, H. Leslie, Y. Heryadi, and E. Abdurachman, "FAKE NEWS DETECTION MODEL BASED ON CREDIBILITY MEASUREMENT FOR INDONESIAN ONLINE NEWS," *J. Theor. Appl. Inf. Technol.*, vol. 15, no. 7, 2021, [Online]. Available: www.jatit.org.