Cosine Similarity untuk Mengukur Tingkat Kesadaran pada Topik *Software Security* Berbasis Teks Komentar di Media Sosial Youtube

Alfirna Rizqi Lahitani [1)], Ulfi Saidata Aesyi [2)], Noviana Wulandari [3)], Bagas Dwi Santosa [4)]

[1)3)] Program Studi Teknologi Informasi, Fakultas Teknik dan Teknologi Informasi, Universitas Jenderal Achmad Yani Yogyakarta
Jl. Siliwangi Ringroad Barat, Banyuraden, Gamping, Sleman, Yogyakarta
[1)] alfirnarizqi@gmail.com
[3)] novianawulandari928@gmail.com

[2)4)] Program Studi Sistem Informasi, Fakultas Teknik dan Teknologi Informasi, Universitas Jenderal Achmad Yani Yogyakarta
Jl. Siliwangi Ringroad Barat, Banyuraden, Gamping, Sleman, Yogyakarta
[2)] ulfiaesyi@gmail.com
[4)] bagasdwisantosa87@gmail.com

Abstrak

Kecenderungan peningkatan celah keamanan siber pada kelemahan perangkat lunak akan mengancam kerahasiaan, integritas, dan keamanan tidak hanya secara infrastruktur tetapi dapat menyerang secara psikologis. Kesadaran akan keamanan siber khususnya pada area CyBOK *Software Security* menjadi perhatian bagi para pengguna, sebagian besar aktivitas siber menggunakan *software*. Kurangnya edukasi akan kesadaran membuat tidak sedikit para pengguna menjadi korban dari celah keamanan siber. Penelitian ini bertujuan melakukan analisis data berbasis teks pada 100 komentar pengguna di sosial media Youtube untuk mengukur tingkat kesadaran terhadap keamanan siber pada topik *software security*. Data teks dikolektif, dibersihkan dan ditransformasi menjadi *term* sehingga siap digunakan untuk proses pembobotan. Pembobotan term menggunakan metode TF-IDF, selanjutnya dilakukan pengukuran derajat kemiripan pada topik *software security* menggunakan *Cosine Similarity*. Hasil analisis divisualisasi dalam bentuk derajat *awareness* yang memodelkan tingkat kesadaran pengguna pada topik *software security*. Hasil pengukuran menunjukkan 98% pengguna pada kategori "kurang *aware*", 2% pengguna pada kategori "cukup *aware*".

Kata kunci: *Cosine Similarity*, TF-IDF, Kesadaran, *Software Security*, Media Sosial

Abstract

The trend of increasing cybersecurity vulnerabilities in software will threaten confidentiality, integrity, and security not only in infrastructure but can attack psychologically. According to the APJI 2020 survey, some people only feel safe when doing activities on the internet, but they don't care that threat can come unnoticed. Then, what about the others. Cyber security awareness, especially in the areas of CyBOK Software Security, currently needs special attention for users. Lack of education on awareness makes many users become victims of cybersecurity loopholes. This study aims to analyze text-based data on 100 user comments on YouTube social media to measure the level of awareness of cybersecurity on the topic of software security. Text data is collected, cleaned, and transformed into terms so that it is ready to be used for the weighting process. The weighting of terms using the TF-IDF method, then measuring the degree of similarity on the topic of software security using Cosine Similarity. The results of the analysis show that 98% of users are in the "less aware" category, 2% of users are in the "fairly aware" category.

Keywords: *Cosine Similarity, TF-IDF, Awareness, Software Security, Social Media*

1. PENDAHULUAN

Pengguna internet terus meningkat diiringi dengan aktivitas yang beragam, data statistik APJII 2020 menampilkan sejumlah perilaku, sebanyak 95,4% masyarakat terhubung dengan internet

 Received: 22 November 2022
Accepted: 9Desember 2022

melalui perangkat pintar dengan durasi aktivitas 8 jam atau lebih. Selanjutnya sebanyak 72.8% masyarakat tidak tahu apakah perangkat yang digunakan pernah terkena virus berdasarkan pengalaman saat terhubung dengan Internet [1]. Kemajuan teknologi informasi dan jaringan serta aktivitas yang masif membawa peluang dan ancaman bagi dunia siber, sehingga terbentuk konsep keamanan siber.

Keamanan siber memiliki kodifikasi yang disebut dengan *Cyber Security Body of Knowledge* (CyBOK) sebagai pemetaan pengetahuan yang berkaitan dengan keamanan siber. CyBOK memiliki 5 kodifikasi area, salah satu area pada CyBOK *Software and Platform Security* yaitu *Software Security* [2]. Aktivitas yang dilakukan dalam siber menggunakan perangkat keras baik yang berdiri sendiri maupun perangkat bergerak di dalamnya melekat sistem operasi dan aplikasi yang merupakan *software*. Sehingga dapat dikatakan aktivitas siber bergantung kepada *software*. Kegagalan *software* dapat berakibat ketidaknyamanan, umumnya berupa aplikasi tidak berfungsi seperti yang diharapkan, namun kegagalan *software* dapat juga menimbulkan masalah pada aspek keamanan *software security* yang meliputi *confidentiality* (kerahasiaan data), *integrity* (integritas data) dan *availability* (ketersediaan).

Jumlah kelemahan perangkat lunak semakin meningkat pada 5 tahun terakhir (2017-2021), khususnya pada tahun 2021 mencapai 18.378 kerentanan yang ditemukan, naik 13% dari tahun sebelumnya [3]. Sebuah portal berita keamanan siber mencatat kasus keamanan *software* terjadi dalam bentuk *bug* atau *error*, sebagai contoh kasus *bug* pada sistem operasi iOS 7 yang digunakan oleh Apple dalam produk iPhone mengandung kesalahan kode mengakibatkan alur logika salah, yang semula ditujukan untuk menguji validasi sertifikat SSL akibatnya banyak pihak dapat masuk ke perangkat dengan sertifikat palsu. Baru-baru ini terjadi kembali pada iOS 15, *bug* dalam sistem mengakibatkan kebocoran aktivitas penjelajahan pengguna secara *real time* [4].

Penyebab kegagalan *software* dapat diidentifikasi dari proses pengembangan *software*, namun *programmer* maupun pengembang yang telah menyadari akan memperbaiki kesalahan sehingga kegagalan *software* dapat teratasi, kepercayaan pengguna untuk memakai *software* dapat dipertahankan. Pada sisi pengguna umumnya percaya dan merasa aman bahwa *software* yang dipilih dapat berjalan sebagaimana mestinya, kegagalan atas *software* yang digunakan diketahui ketika terlihat dan berdampak langsung. Berdasarkan penggunaan dan dampaknya, apakah pengguna menyadari pentingnya pengetahuan tentang perangkat yang digunakan serta hal-hal yang berkaitan dengan keamanan *software*.

Untuk mengetahui bagaimana kondisi literasi digital saat ini, pengukuran tingkat kesadaran dan pengetahuan tentang *cybersecurity* telah dilakukan pada penelitian sebelumnya dengan beragam metode [5][6][7][8]. Secara umum, metode yang digunakan adalah *Technology Acceptance Model* (TAM), *Structural Equation Modeling* (SEM) serta metode kuantitatif menggunakan kuesioner. Hasil penelitian menunjukkan tingkat kesadaran terhadap keamanan siber berada pada tingkat sedang. Pengambilan data seluruhnya pada pengguna dengan tingkat partisipasi aktif di media sosial, dari hasil penelitian ditemukan bahwa pengguna yang pernah mengalami ancaman pada akun media sosial lebih menunjukkan kesadaran mereka dengan meningkatkan level keamanan *password* akun media sosial, dan lebih waspada sebelum mengijinkan akses ke perangkat pribadi pengguna dari akun media sosial. Oleh karena itu, banyak langkah proaktif yang perlu dilakukan oleh para pemangku kepentingan agar isu-isu yang relevan dengan *cybercrime* dapat dikurangi dan kemudian dihilangkan.

Penelitian ini dilakukan menjadi bagian dari peran keamanan siber dalam memastikan dan menemukan pengetahuan pada aktivitas siber melalui pengukuran level of *awareness* pada topik *software security*. Perbedaan pada penelitian sebelumnya adalah metode yang dilakukan untuk mengukur tingkat kesadaran terhadap literasi digital *cybersecurity* pada media sosial umumnya dilakukan menggunakan metode survei, kuesioner, dan metode statistik analisis.

Pada penelitian ini pengukuran tingkat kesadaran dilakukan dengan teknik yang berbeda, dengan mengamati aktivitas siber pada topik *software security* dan mengumpulkan data komentar berbasis teks. Pengukuran kesadaran hanya dilakukan pada dimensi pengetahuan, tidak mengukur sikap dan perilaku. Pada objek data teks media sosial tidak memungkinkan mengukur dimensi sikap dan perilaku dikarenakan tidak melakukan pengamatan langsung pada objek personal.

Sedangkan objek yang dianalisis pada Penelitian ini merupakan objek dengan tipe data teks yang membutuhkan proses ekstraksi, sehingga metode yang tepat digunakan adalah metode *text mining* untuk menemukan dan memetakan level of *awareness*.

Penelitian sebelumnya yang berfokus pada *data mining* khususnya teknik pengukuran kemiripan *term* [9][10][11][12]. Teknik ini akan diimplementasikan untuk analisis keamanan siber. Kesadaran akan keamanan siber khususnya pada area *Software Security* perlu menjadi perhatian bagi para pengguna. Kurangnya edukasi akan kesadaran membuat tidak sedikit para pengguna menjadi korban dari celah keamanan siber.

Penelitian ini bertujuan melakukan analisis data berbasis teks pada komentar di sosial media youtube untuk mengukur tingkat kesadaran pengguna terhadap keamanan siber pada topik *software security*. Youtube dipilih karena 61% pengguna sering menonton Youtube dan salah satu media sosial yang menyajikan konten dalam bentuk audio visual serta memiliki interaksi dalam bentuk komentar. Data teks dikolektif, dibersihkan dan ditransformasi menjadi term sehingga siap digunakan untuk proses pembobotan. Pembobotan term menggunakan metode TF-IDF, selanjutnya dilakukan pengukuran derajat kemiripan pada topik *software security* menggunakan *Cosine Similarity*. Hasil analisis divisualisasi dalam bentuk derajat *awareness* yang memodelkan tingkat kesadaran pengguna pada topik *software security*.

2. TINJAUAN PUSTAKA

2.1 *Cybersecurity*

Cybersecurtity atau Keamanan siber merujuk pada konsep perlindungan keamanan meliputi alat, strategi, pedoman, metode manajemen resiko, jaminan mencakup perangkat keras, perangkat lunak, data dan layanan untuk melindungi lingkungan jaringan, aset organisasi dan pengguna dari ancaman bahaya, penyalahgunaan dan akses yang tidak sah [13].

Keamanan siber memiliki kodifikasi yang disebut dengan *Cybersecurity Body of Knowledge* (CyBOK) sebagai pemetaan pengetahuan yang berkaitan dengan keamanan siber. Pengetahuan yang dikodifikasi telah ada dalam buku, artikel, penelitian, laporan dan standar akademik yang dapat digunakan sebagai pedoman, standar pendidikan maupun dikembangkan berdasarkan CyBOK. CyBOK memiliki 5 kodifikasi area diantaranya adalah (1) *Human, Organisational, and Regulatory Aspects*, (2) *Attacks and Defences*, (3) *Systems Security*, (4) *Software and Platform Security* dan (5) *Infrastructure Security*.

Adapun peran keamanan siber yaitu menemukan, memperbaiki atau mengurangi resiko terjadinya kejahatan (*cybercrime*) termasuk ancaman (*cyberthreat*) dan serangan (*cyberattack*) dari seluruh aktivitas teknologi siber [14].

2.2 *Awareness*

Awareness (kesadaran) erat berkaitan dengan pengetahuan. Kesadaran bermakna bahwa seseorang menyadari dalam arti mengetahui stimulus (objek) terlebih dahulu. Untuk menemukan sejauh mana kesadaran, diperlukan pengukuran tingkat terhadap objek, menurut model yang dikembangkan oleh Kruger dan Kearney kesadaran diukur dalam 3 dimensi yaitu pengetahuan, sikap dan perilaku, masing-masing dimensi memiliki proporsi yang berbeda, *knowledge* (pengetahuan) 30%, *attitude* (sikap) 20%, *behaviour* (perilaku) 50% [15]. Pengetahuan adalah hasil yang terjadi setelah orang melakukan penginderaan terhadap suatu objek tertentu. Penginderaan ini terjadi melalui panca indera manusia, yaitu penglihatan, pendengaran, penciuman rasa dan raba. Pengetahuan atau kognitif merupakan domain yang sangat penting dalam membentuk tindakan seseorang, aspek pengetahuan akan menentukan sikap.

2.3 *Data Mining*

Data mining adalah bidang yang mempelajari tentang teknik pengumpulan, pembersihan, proses, analisis data sehingga menghasilkan sesuatu yang dapat menambah wawasan[16].

Data mining sebagai proses menemukan struktur yang menarik dalam data, Sesi penambangan data menggunakan satu atau beberapa algoritme untuk tujuan identifikasi tren dan pola menarik dalam data[17].

p-ISSN: 2460-173X
e-ISSN: 2598-5841

Tahap *data mining* meliputi *(1) Defining the problem*, beberapa hal yang dilakukan adalah mendefinisikan masalah dan tujuan, identifikasi kebutuhan dan menentukan teknik yang tepat untuk diimplementasikan pada setiap tahap analisis. *(2) Data Preparation*, merupakan tahap pengumpulan data yang dibutuhkan untuk selanjutnya mempersiapkan data sebelum dilakukan pemrosesan dan analisis, data berbasis teks akan melalui *preprocessing* dan disajikan dalam bentuk *final dataset* yang siap untuk dilakukan perhitungan. *(3) Data Exploration*, merupakan tahap seleksi teknik pemrosesan dan analisis data yang sesuai dengan *dataset* yang dimiliki. *(4) Modeling,* menerapkan teknik, metode dan algoritma *Data Mining*. *(5) Evaluating and Validating the Model,* tahap ini melibatkan proses evaluasi data *training* dan *testing*. *(6) Deploy and Updation,* tahap penarikan keputusan. [18].

2.4 *Term Frequency – Invers Document Frequency* (TF-TDF)

Metode TF-IDF untuk mendapatkan bobot perwakilan dari kata-kata yang diekstrak dari data informasi dengan mempertimbangan penyebaran kata di dokumen lain. Formula TF-IDF ditunjukkan seperti pada persamaan (1) dan (2) [19].

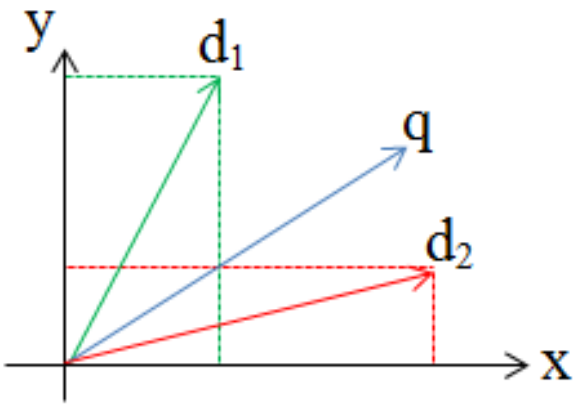
$$\boldsymbol{idf_i = \log\left(\frac{N}{df_i}\right)} \tag{1}$$

$$\boldsymbol{w_{i,j} = tf_{i,j} \times idf_i} \tag{2}$$

Keterangan:
w = bobot *term* pada dokumen (i,j: 1,2,3,…n)
n = total keseluruhan dokumen yang ada
t = *term* atau kata yang akan dihitung bobotnya. (i: 1,2,3,…n)
d = dokumen yang mengandung sekumpulan *term* (j: 1,2,3,...n)
df: frekuensi atau jumlah dokumen yang berisi *term* (i: 1,2,3,…n)
tf = frekuensi atau jumlah *term* yang muncul dalam sebuah dokumen (i,j: 1,2,3,...n)
idf = *inverse* frekuensi dokumen, nilai kemunculan *term* pada kumpulan dokumen (i: 1,2,3,…n)

2.5 *Cosine Similarity*

Cosine Similarity merupakan perhitungan untuk mengukur tingkat kemiripan yang didapatkan dari nilai *cosinus* sudut perkalian dua buah vektor yang dibandingkan, karena nilai *cosinus* 0˚ adalah 1 dan kurang dari 1 untuk nilai sudut yang lain, maka nilai similaritas dari dua vektor yang dikatakan mirip ketika nilai dari *Cosine Similarity*. Ilustrasi kerapatan sudut kosinus seperti yang ditunjukkan pada Gambar 1.[20].



Gambar 1. Ilustrasi kerapatan sudut kosinus

Perhitungan *Cosine Similarity* dilakukan dengan Persamaan (3):

p-ISSN: 2460-173X
e-ISSN: 2598-5841

$$Cos\ \alpha = \frac{A \times B}{|A| \times |B|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} (A_i)^2} \times \sqrt{\sum_{i=1}^{n} (B_i)^2}} \quad (3)$$

Keterangan:
n = jumlah dokumen yang diuji (i:1,2,3..n)
A = bobot *term* pada dokumen *query* (ahli) (i: 1,2,3..n)
B = bobot *term* dokumen uji (siswa) (i:1,2,3..n)
|A| = panjang vektor dokumen A
|B| = panjang vektor dokumen B

3. METODE PENELITIAN

Penelitian ini menggunakan metode *data mining* untuk menemukan Informasi dengan menggunakan sejumlah teknik, algoritma maupun perhitungan matematika dalam menemukan pola atau model yang terdapat pada data. Adapun alur tahapan *data mining* meliputi[18]:

1. *Defining the problem*, beberapa hal yang dilakukan adalah mendefinisikan masalah dan tujuan, identifikasi kebutuhan dan menentukan teknik yang tepat untuk diimplementasikan pada setiap tahap analisis.
2. *Data Preparation*, merupakan tahap pengumpulan data yang dibutuhkan untuk selanjutnya mempersiapkan data sebelum dilakukan pemrosesan dan analisis, data berbasis teks akan melalui preprocessing dan disajikan dalam bentuk final dataset yang siap untuk dilakukan perhitungan.
3. *Data Exploration*, merupakan tahap seleksi teknik pemrosesan dan analisis data yang sesuai dengan dataset yang dimiliki.
4. *Modeling*, menerapkan teknik, metode dan algoritma Data Mining.
5. *Evaluating and Validating the Model*, tahap ini melibatkan proses evaluasi data training dan testing.
6. *Deploy and Updation*, tahap penarikan keputusan.

Data yang menjadi objek pada Penelitian ini adalah kumpulan komentar pada sebuah channel sosial media Youtube yang merupakan data berbasis teks, eksplorasi yang dilakukan pada sejumlah data teks untuk memperoleh Informasi tertentu masuk pada ranah *text mining* sehingga perlu ada *preprocessing* terlebih dahulu untuk mempersiapkan data agar siap diolah. Teknik yang digunakan untuk menentukan *awareness* adalah pengukuran similaritas menggunakan metode *Cosine Similarity* dan TF-IDF. Pada setiap tahap analisis dilakukan dengan memanfaatkan *tools Scrapping* dan *programming* Python untuk mempermudah dalam eksekusi data.

4. PEMBAHASAN

Pada penelitian ini pengukuran tingkat kesadaran berasal dari data komentar berbasis teks pada media sosial Youtube. Topik keamanan siber yang ditentukan adalah *software security*. Dipilih sebuah *chanel* yang akan diolah untuk menemukan *term* kunci yang disebut sebagai dokumen sumber (Q), dan sebuah chanel yang akan diambil datanya yang disebut sebagai dokumen uji (K).

Pengumpulan data dilakukan menggunakan teknik *scrapping* dengan tujuan mengambil data tertentu pada sebuah halaman *web*. Adapun data yang dikumpulkan adalah data komentar dari *chanel* Youtube "Keamanan Informasi: *Software Security*" pada alamat: https://www.youtube.com/watch?v=UiiY5AX5uak .

Hasil *scrapping* akan dipecah menjadi 2 dokumen yaitu dokumen sumber (Q) yang merupakan dokumen sumber yang menjadi pembanding pada proses pengukuran similaritas, dan dokumen uji (Kn) yang merupakan dokumen uji terdiri dari setiap komentar dari pengguna. Data yang diperoleh dijabarkan pada Tabel 1.

Tabel 1. Informasi hasil pengumpulan data

No	Item	Keterangan
1	Dokumen Sumber (Q)	Merupakan dokumen sumber yang menjadi pembanding pada proses pengukuran similaritas; Hasil Scrapping 149 baris; Hasil preprocess 68 terms kunci; Telah mewakili konten/materi pembahasan pada topik *Software Security*.
2	Dokumen Uji (K_n)	Merupakan dokumen uji yang terdiri dari setiap komentar dari pengguna. Hasil Scrapping 100 baris; Hasil preprocessing langsung ditransformasi dalam indeks *term*; Masing-masing komentar akan dianalisis kemiripannya.

Selanjutnya hasil *scrapping* yang tersimpan dikonversi dalam bentuk ".csv" untuk dilakukan *preprocessing* menggunakan Phyton. Adapun *preprocessing* terdiri dari tahap-tahap sebagai berikut: (1) *Case Folding*, (2) *Tokenizing*, (3) *Stopword Removal*, (4) *Stemming* dan (5) *Indexing*.

Preprocessing yang dilakukan pada dokumen sumber akan menjadi indeks *term* kunci pada dokumen (Q) yang terdiri dari 68 *terms*, sedangkan *preprocessing* 100 dokumen uji akan dihitung dulu jumlah kemunculan *term* kunci yang terkandung pada masing-masing dokumen uji (Kn). Untuk memudahkan pengamatan, berikut adalah cuplikan hasil *preprocessing* data dijabarkan pada Tabel 2.

Tabel 2. Hasil *Preprocessing*

['ada', 'command', 'terminal', 'chmod', '777', 'file', 'directory', 'contoh', 'acl', 'linux', 'bapa', '4753']
['izin', '1631', 'pasang', 'mendownload', 'plugin', 'baja', 'dalam', 'malware', 'rugi', 'web', 'rugi', 'perangkat', 'web', 'orang', 'serang', 'web', 'pribadi', 'plugin', 'baja', 'jadi', 'zombie', 'ddos', 'rugi', 'web', 'orang']
['934', 'tambah', 'web', 'aplikasi', 'sukses', 'serang', 'contoh', 'orang', 'rela', 'informasi', 'orang', 'tanggung', 'contoh', 'super', 'user', 'idpassword', 'orang']
['menit', '2317', 'information', 'security', 'threat', 'ijin', 'tindak', 'skimming', 'dalam', 'ancam', 'ganggu', 'aman', 'infromasi', 'skimming', 'salah', 'ganggu', 'aman', 'infromasi', 'aspek', 'ancam', 'interruption', 'interception', 'modification', 'fabrication', 'terimakasih']
['pa', 'interception', '2320', 'sadap', 'australia', 'indonesia']
['kait', 'bahas', 'aman', 'informasi', '2322', 'whatapp', 'web', 'ganggu', 'aman', 'informasi', 'bgian', 'threat']
['ijin', 'access', 'control', 'models', 'bapa', '4855', 'model', 'aplikasi', '1', 'buah', 'akses', 'model']

Term yang telah melalui *preprocessing* diindeks dan dimodelkan sebagai dokumen kunci untuk selanjutnya dibandingkan dengan dokumen uji dengan cara menghitung jumlah kemunculan term kunci pada dokumen uji. Berikut adalah cuplikan hasil pengolahan data pada 1-10 data uji yang dijabarkan pada Tabel 3.

Tabel 3. Hasil pembobotan *term* dan pengukuran kemiripan

Term	Q	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K100	df	idf
command	1	1											5	1.30103
terminal	1	1											4	1.39794
file	1	1											6	1.221849
directory	1	1											3	1.522879
linux	1	1											5	1.30103
plugin	1		2										3	1.522879
malware	1		1										6	1.221849

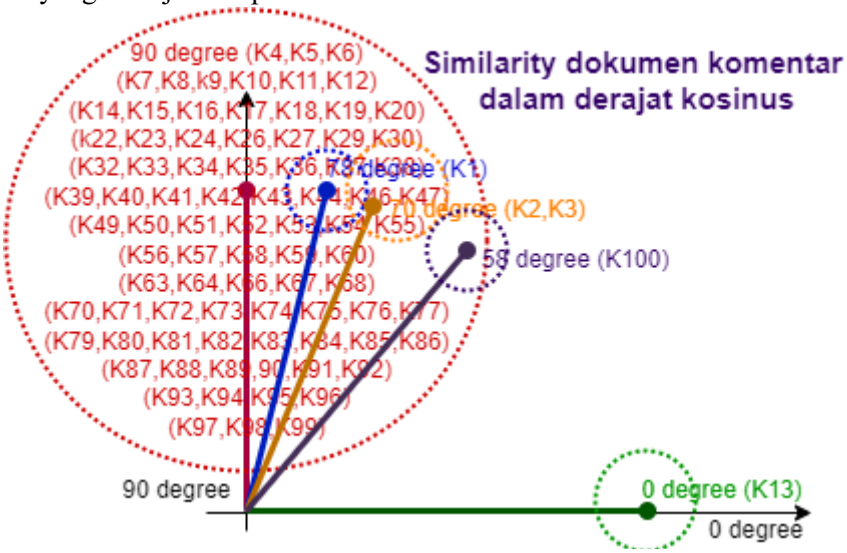
p-ISSN: 2460-173X
e-ISSN: 2598-5841

Term	Q	K1	K2	K3	K4	K5	K6	K7	K8	K9	K10	K100	df	idf
web	5		4	1			1						20	0.69897
perangkat	2		1										6	1.221849
aplikasi	4			1				1					18	0.744727
Pvektor	16	3	4.5	2.2	6	2	3	6	4	3	2	9.98		
Cosine Similarity		0.2	0.3	0.3	0	0	0	0	0	0	0	0.52		

Hasil pengukuran similaritas dibagi dalam 3 kategori diantaranya (1) Kategori Baik dengan nilai *Cosine Similarity* antara 0.8-1, (2) kategori Cukup dengan nilai *Cosine Similarity* 0.5-0.8, (3) Kategori Kurang dengan nilai *Cosine Similarity* antara 0-0.5. Hasil Pengukuran *Awareness* dalam 3 Kategori serta jumlah komentar ditunjukkan pada Gambar 2.

Gambar 2. Visualisasi hasil pengukuran berdasarkan kategori *level of Awareness*

Hasil pengukuran similaritas masing-masing komentar divisualisasikan dalam bentuk derajat kosinus seperti yang ditunjukkan pada Gambar 3.



Gambar 3. Visualisasi hasil pengukuran *similarity* berdasarkan kerapatan derajat kosinus

5. KESIMPULAN

Berdasarkan hasil analisis divisualisasi dalam bentuk derajat *awareness* yang memodelkan tingkat kesadaran pengguna pada topik *software security*. Derajat awareness merupakan representasi nilai similaritas dari derajat kosinus. Nilai similaritas didapatkan dari pertemuan sudut antara vektor r dan vektor x, kisaran nilai similaritas adalah 0 hingga 1 dimana dua vektor dikatakan mirip ketika nilai sudut cosinus $0^o = 1$. Semakin rapat jarak sudut cosinus kedua vektor maka semakin tinggi nilai similaritas antar vektor tersebut. Hasil pengukuran menunjukkan 98% pengguna pada kategori "kurang *aware*", 2% pengguna pada kategori "cukup *aware*". Penelitian ini dilakukan menjadi bagian dari peran keamanan siber dalam memastikan dan menemukan pengetahuan pada aktivitas siber melalui pengukuran *level of awareness* pada topik *software security* dan memberikan kontribusi 30% dalam menemukan hasil pengukuran *awareness* pada dimensi pengetahuan.

UCAPAN TERIMAKASIH

Penulis mengucap syukur kepada Tuhan Yang Maha Esa, karena atas rahmat dan berkat-Nya, penulis dapat menyelesaikan artikel dan penelitian ini dengan tepat waktu. Ucapan terimakasih ini ditujukan kepada Direktorat Riset, Teknologi, dan Pengabdian kepada Masyarakat (DRTPM) Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi (Kemdikbudristek) Republik Indonesia yang telah memberikan kesempatan kepada tim peneliti untuk melakukan penelitian pada tahun 2022 pada skema Penelitian Dosen Pemula (PDP).

DAFTAR PUSTAKA

[1] APJII, "Laporan Survei Internet APJII 2019 – 2020," *Asosiasi Penyelenggara Jasa Internet Indonesia.*, vol. 2020, pp. 1–146, 2020, [Online]. Available: https://apjii.or.id/survei.

[2] A. Martin and H. Chivers, "Introduction to CyBOK," 2019.

[3] J. Greig, "With 18,378 vulnerabilities reported in 2021, NIST records fifth straight year of record numbers," 2021. https://www.zdnet.com/article/with-18376-vulnerabilities-found-in-2021-nist-reports-fifth-straight-year-of-record-numbers/.

[4] A. T. Dan Goodin, "A Bug in iOS 15 Is Leaking User Browsing Activity in Real Time," 2022. https://www.wired.com/story/ios-15-bug-leaking-user-browsing-activity-in-real-time/.

[5] D. Revilia and Irwansyah, "Literasi Media Sosial: Kesadaran Keamanan Dan Privasi Dalam Perspektif Generasi Milenial Social," *J. Penelit. Komun. dan Opini Publik*, vol. 24, no. 1, pp. 1–15, 2020.

[6] R. Riyandhika and R. Pratama, "Analisis Kesadaran Cybersecurity pada Kalangan Mahasiswa di Indonesia," *Uii*, vol. 1, no. 2, p. 1, 2020.

[7] M. R. Ramadhani and A. R. Pratama, "Analisis Kesadaran Cybersecurity Pada Pengguna Media Sosial Di Indonesia," *Journal.Uii.Ac.Id*, vol. 1, no. 2, pp. 1–8, 2020.

[8] S. H. Et. al., "Level of Awareness of Social Media Users on Cyber Security: Case Study among Students of University Tun Hussein Onn Malaysia," *Turkish J. Comput. Math. Educ.*, vol. 12, no. 2, pp. 694–698, 2021, doi: 10.17762/turcomat.v12i2.923.

[9] C. De Boom, S. Van Canneyt, S. Bohez, T. Demeester, and B. Dhoedt, "Learning Semantic Similarity for Very Short Texts," 2015, doi: 10.1109/ICDMW.2015.86.

[10] A. L. Buczak and E. Guven, "A Survey of Data Mining and Machine Learning Methods for Cyber Security Intrusion Detection," *IEEE Commun. Surv. Tutorials*, vol. 18, no. 2, pp. 1153–1176, 2016, doi: 10.1109/COMST.2015.2494502.

[11] A. R. Lahitani, A. E. Permanasari, and N. A. Setiawan, "Cosine similarity to determine similarity measure: Study case in online essay assessment," 2016, doi: 10.1109/CITSM.2016.7577578.

[12] A. R. Lahitani, "Automated Essay Scoring menggunakan Cosine Similarity pada Penilaian Esai Multi Soal," *J. Kaji. Ilm.*, vol. 22, no. 2, pp. 107–118, 2022, doi: 10.31599/jki.v22i2.1121.

[13] H. S. D. N. dan A. G. T. Pitra Ratulangi, "Jenis Kejahatan Pada Masa Pandemi Covid-19 Dalam Perspektif Cyber Security Nasional Di Indonesia," vol. 3, no. March, p. 6, 2021.
[14] L. Siagian, A. Budiarto, P. Strategi, P. Udara, and U. Pertahanan, "Peran Keamanan Siber Dalam Mengatasi Konten Negatif Guna Mewujudkan Ketahanan Informasi Nasional," pp. 1–18, 2017.
[15] B. Ngoqo and S. V. Flowerday, "Information Security Behaviour Profiling Framework (ISBPF) for student mobile phone users," *Comput. Secur.*, vol. 53, no. July 2016, pp. 132–142, 2015, doi: 10.1016/j.cose.2015.05.011.
[16] C. Aggarwal, "An Introduction to Data Mining," in *Data Mining*, 2015.
[17] R. J. Roiger, *Data Mining: A Tutorial-Based Primer, Second Edition*. 2017.
[18] B. M. B. Rahul D. Shanbhogue, "Survey of Data Mining (DM) and Machine Learning (ML) Methods on Cyber Security," *Indian J. Sci. Technol.*, vol. 10, no. 35, pp. 1–7, 2017, doi: 10.17485/ijst/2017/v10i35/118951.
[19] J. T. Elektro and F. Teknik, *Implementasi Metode Maximum Marginal Relevance Pada Peringkasan Teks*. 2015.
[20] P. Informasi, "Penelusuran Informasi (Information Retrieval) - Pembobotan Kata dan Vector Space."

Biodata Penulis

Alfirna Rizqi Lahitani, Lahir di Sintang pada 06 Januari 1992. Lulus Pendidikan S1 pada tahun 2013 Informatika STMIK Jenderal Achmad Yani Yogyakarta. Pendidikan S2 tahun 2017 Magister Teknologi Informasi Universitas Gadjah Mada. Saat ini merupakan Dosen Program Studi Teknologi Informasi Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta.

Ulfi Saidata Aesyi, Lahir di Magelang pada 15 Desember 1990. Lulus pendidikan S1 pada tahun 2013 Teknik Informatika Universitas Muhammadiyah Magelang. Lulus Pendidikan S2 tahun 2017 Ilmu Komputer Univesitas Gadjah Mada. Saat ini Merupakan Dosen Program Studi Sistem Informasi Fakultas Teknik dan Teknologi Informasi Universitas Jenderal Achmad Yani Yogyakarta.

Noviana Wulandari, lahir di Trenggalek Jawa Timur 2002. Menempuh pendidikan di SMK Negeri 2 Trenggalek dan melanjutkan studi S1 di program studi Teknologi Inoformasi Universitas Jendral Acmad Yani Yogyakarta. Selain menjadi mahasiswa, juga bekerja paruh waktu yang dilakukan setelah pulang kuliah.

Bagas Dwi Santosa, pernah bersekolah di SDN 1 Ngerangan (2009 – 2015), SMPN 1 Bayat (2015-2018), SMKN 1 Ngawen (2019-2021), Sekarang, tengah menempuh studi strata satu semester tiga di Universitas Jenderal Achmad Yani Yogyakarta Fakultas Teknik dan Teknologi Informasi, mengambil program studi Sistem Informasi. Kesibukan penulis sebagai mahasiswa aktif dan pengalaman organisasi sebagai staff Research and Development Himpunan Mahasiswa, Core team Google Developer Student Club, serta kepanitiaan dibeberapa acara kampus.