

METODE VECTOR SPACE MODEL UNTUK WEB SCRAPING PADA WEBSITE FREELANCE

Andi Nurkholis^{1*}; Yusra Fernando^{2*}; Faris Arkans Ans³

Informatika^{1,2,3}

Universitas Teknokrat Indonesia

www.teknokrat.ac.id

andinh@teknokrat.ac.id^{1*}; yusra.fernando@teknokrat.ac.id^{2*}; faris.arkan.ans@gmail.com³



Abstract— In digitalization era, internet is at the center of all lines of community activity, just like the field of work. Currently, many platforms provide job vacancies, especially for freelancers. To obtain this information, users usually need to open several websites to find information about suitable job vacancies. Web scraping offers solution to overcome these problems. Based on research that has been done, the BeautifulSoup and Selenium libraries will be used to collect data. To search for data, vector space model method is used to find the level of data similarity between the query and the document. In exploring data, the average near-perfect recall value is 98%, while the average precision value is 56%. This is because data search uses three parameters, so the possibility of retrieving irrelevant data is more significant if the document contains a word in the user's query, even though the context does not match. Utilizing the Streamlit framework in Python can display the data processing results and help users navigate the web scraping process, data processing, and data search. This study aims to implement the web scraping method to retrieve data from freelance websites: Freelance, Project, and Sribulancer. By applying the vector space model method, users can search data from several websites without opening freelance websites one by one. Using data visualization in the form of a web application using the Streamlit framework, the web scraping results can also be processed to be presented in a more helpful form and save the user's time.

Keywords: data visualization, freelance job, vector space model, web scraping

Abstrak— Di zaman digitalisasi saat ini, internet menjadi pusat dari semua lini kegiatan masyarakat, seperti halnya bidang pekerjaan. Saat ini sudah banyak platform yang menyediakan lowongan pekerjaan, khususnya bagi para pekerja lepas. Untuk memperoleh informasi tersebut, biasanya pengguna perlu membuka beberapa website untuk menemukan informasi tentang lowongan kerja yang sesuai. Web scraping menawarkan sebuah solusi untuk mengatasi permasalahan tersebut. Berdasarkan penelitian yang telah dilakukan sebelumnya untuk pengambilan data akan digunakan library BeautifulSoup dan Selenium. Untuk pencarian data digunakan metode vector space model untuk mencari tingkat kemiripan data Antara query dan dokumen. Dalam pencarian data dihasilkan rata-rata nilai recall yang mendekati sempurna yaitu sebesar 98%, sedangkan untuk rata-rata nilai precision sebesar 56%. Hal tersebut dikarenakan pencarian data menggunakan 3 parameter sehingga kemungkinan untuk mengambil data yang tidak relevan juga semakin besar jika dokumen tersebut mengandung sebuah kata pada query user walaupun konteksnya tidak sesuai. Dengan memanfaatkan framework streamlit pada Python hasil proses dari data tersebut dapat ditampilkan serta membantu pengguna untuk menavigasi proses web scraping, proses data, dan pencarian data. Penelitian ini bertujuan untuk mengimplementasikan metode web scraping untuk mengambil data dari beberapa website freelance yaitu :Freelance, Project, dan Sribulancer, lalu dengan menerapkan metode vector space model pengguna dapat melakukan pencarian data dari beberapa website tanpa harus membuka website freelance satu per satu. Menggunakan visualisasi data berupa aplikasi web menggunakan framework streamlit, hasil dari web scraping tersebut juga dapat diproses untuk selanjutnya disajikan dalam bentuk yang lebih bermanfaat dan menghemat waktu dari pengguna.

Kata kunci: pekerja lepas, vector space model, visualisasi data, web scraping

PENDAHULUAN

Perkembangan teknologi informasi kian maju setiap waktunya. Segala sesuatunya telah masuk pada era digitalisasi berkat adanya jaringan

internet sebagai media komunikasi data. Internet telah menjadi pusat dalam kegiatan masyarakat, tidak terkecuali pada bidang pekerjaan. Pada umumnya, pekerjaan atau sebuah bisnis sangat memerlukan tempat/lokasi kantor serta memiliki

pekerja/karyawan tetap. Namun, hal tersebut kini bukan lagi menjadi urgensi, dikarenakan situs freelance dapat dimanfaatkan sebagai platform yang mampu menghubungkan pekerja dan pemilik bisnis. Tanpa adanya kantor, para pekerja memiliki fleksibilitas untuk menjalankan tugas mereka di berbagai tempat dan waktu yang sesuai, asalkan mereka tetap terhubung melalui internet. (Firrezqi, 2020). Profesi tersebut sering disebut sebagai pekerja lepas (freelancer) (Ayoobzadeh, 2022) yang akhir-akhir ini sedang banyak diminati oleh pekerja profesional yang biasa ditemukan pada situs web Freelance, Project, dan Sribulancer. Indonesia menempati peringkat ke-16 dalam daftar negara dengan perekonomian terbesar, dengan jumlah pekerja profesional sebanyak 55 juta orang. Namun, pada tahun 2030, perkiraan jumlah tersebut diperkirakan akan meningkat menjadi 113 juta orang. Dengan pertumbuhan yang pesat ini, Indonesia diproyeksikan akan naik peringkat dan menjadi negara dengan perekonomian terbesar ke-7 di dunia (Mustofa, 2018).

Saat ini sudah banyak platform pihak ketiga yang khusus menyediakan lowongan bagi para freelancer. Permasalahan muncul saat pengguna perlu membuka banyak website untuk menemukan informasi tentang lowongan kerja yang sesuai (Kadam et al., 2018). Hal tersebut dikarenakan tidak sedikit dari para penyedia lowongan freelance menginformasikan lowongan mereka hanya pada satu situs web. Sehingga, banyaknya website freelance akan mengakibatkan sulitnya para freelancer untuk melakukan summarisasi keseluruhan informasi lowongan. Permasalahan tersebut dapat diselesaikan menggunakan teknik web scraping yang bertujuan mengekstrak kerangka halaman website untuk keperluan mengumpulkan data (Henrys, 2021; Thomas & Mathur, 2019) yang dapat diolah menjadi rangkuman informasi dalam waktu singkat (Eminagaoglu, 2022; Ridwan & Hermawan, 2019).

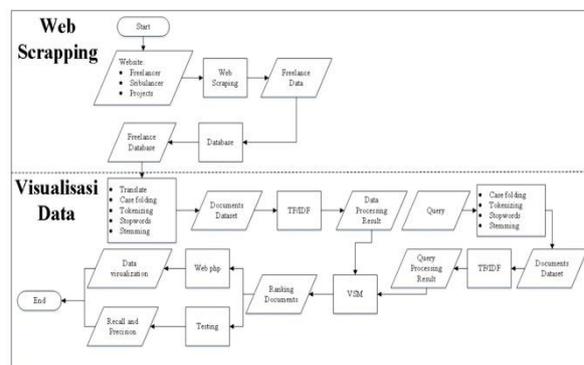
Penelitian yang mengkaji terkait implementasi web scraping telah banyak dilakukan. Penelitian pertama mengimplementasi Web Scraping pada Situs Perhotelan yang berhasil memberikan sampel data dalam jumlah besar secara otomatis untuk selanjutnya dilakukan analisis (Han & Anderson, 2021). Penelitian lainnya menerapkan teknik crawling menggunakan Selenium yang secara otomatis melakukan navigasi dan perintah masukan user secara otomatis pada halaman web (Semendawai et al., 2021). Pada penelitian lainnya, hasil implementasi web scraping juga dapat ditindaklanjuti dengan penerapan metode vector space model (VSM) untuk visualisasi informasi lagu karaoke berdasarkan pencarian pengguna yang dikombinasikan pengurutan berdasarkan tingkat kemiripan kata

(Anna & Hendini, 2018). VSM digunakan untuk memeriksa tingkat kedekatan atau kesamaan antara istilah-istilah dengan mempertimbangkan pembobotan yang diberikan pada setiap istilah tersebut. (Belwal et al., 2021; Singh & Singh, 2021). Berdasarkan penelitian sebelumnya, teknik web scrapping yang dikombinasikan dengan metode VSM mampu melakukan ekstraksi data yang kemudian divisualisasikan sesuai kebutuhan pengguna.

Penelitian ini bertujuan mengimplementasikan teknik web scraping untuk ekstraksi data pada tiga situs web freelance, yakni Freelance, Project, dan Sribulancer. Selain itu, metode VSM juga digunakan untuk visualisasi hasil pada aplikasi berbasis web dengan memanfaatkan framework streamlit. Untuk menyelesaikan penelitian yang dilakukan, dibagi menjadi dua tahap utama, yakni web scraping dan visualisasi data. Sebagai implikasi, rangkuman dari keseluruhan informasi lowongan kerja freelance dapat ditampilkan secara efisien sesuai kebutuhan pengguna, sehingga akan lebih menghemat waktu.

BAHAN DAN METODE

Objek penelitian ini terdiri dari tiga situs web freelance terbesar di Indonesia, yakni Freelance, Project, dan Sribulancer. Hal tersebut dibuktikan dengan jumlah traffic terbesar dari beberapa situs web freelance lainnya yang diakses oleh masyarakat Indonesia. Tahapan utama penelitian ini dibagi menjadi 2, yaitu web scraping dan visualisasi data. Tahapan web scraping terdiri dari proses Pengumpulan Data dan Penyimpanan Data. Sedangkan proses visualisasi data terdiri dari Preprocessing Data dan Query, Pembobotan TF-IDF, Perangkingan Dokumen dengan Vector Space Model dan Visualisasi Data. Berikut adalah tahapan penelitian yang dapat dilihat pada Gambar 1.



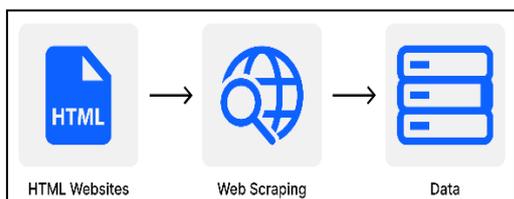
Sumber: (Nurkholis et al., 2023)

Gambar 1. Tahapan penelitian

Berdasarkan Gambar 1, berikut adalah penjelasan dari setiap tahapannya:

A. Web Scrapping

Berikut adalah rincian tahapan yang dilakukan pada tahap web scrapping, ditunjukkan pada Gambar 2.



Sumber: (Khder, 2021)

Gambar 2. Tahapan web scrapping

1. Web Scrapping

Web scrapping adalah proses yang dilakukan untuk mengambil data khusus dari sebuah halaman website secara semi-terstruktur (Julianto et al., 2022). Dalam penelitian ini, tahap web scrapping lowongan kerja freelance dilakukan pada tiga website menggunakan library BeautifulSoup dan Selenium pada bahasa pemrograman Python versi 3.9.

2. Data Collecting

Selanjutnya adalah tahapan penyimpanan data hasil web scrapping pada basis data. Dalam penelitian ini, data disimpan dalam basis data SQLite versi 3.39.4.

B. Visualisasi Data

Tahap ini bertujuan memproses data hasil web scrapping dan memvisualisasikan hasil summarisasi menggunakan metode vector space model. Tahapan visualisasi data terbagi menjadi empat tahap, yang dijelaskan sebagai berikut:

1. Preprocessing Data

Proses preprocessing data bertujuan untuk mempersiapkan data mentah atau raw data (Nurkholis et al., 2022) yang dikumpulkan dari berbagai sumber menjadi informasi yang lebih bersih dan siap digunakan untuk tahapan pengolahan selanjutnya. Tahap ini sering disebut sebagai inisiasi awal untuk mengumpulkan seluruh informasi yang tersedia dengan membersihkan, menyeleksi, dan menggabungkan data-data tersebut. Hasil pengumpulan data belum bisa langsung dilakukan pemodelan dikarenakan dalam data tersebut masih terdapat banyak simbol dan kata-kata yang tidak diperlukan, sehingga diperlukan preprocessing data agar data lebih terstruktur dan bersih sehingga dapat diklasifikasikan (Styawati et al., 2021). Adapun, langkah-langkah data preprocessing yang dilakukan mencakup: translate, case

folding, tokenizing, stopwords, dan stemming.

- a. Translate merupakan langkah proses pengalihan yang bertujuan mengubah teks tulis bahasa sumber menjadi teks bahasa sasaran. Dalam penelitian ini, yang diterjemahkan adalah teks berbahasa inggris menjadi bahasa indonesia yang merupakan tujuan akhir dari visualisasi informasi. Hal tersebut dilakukan dengan alasan telah banyak data pekerjaan freelance dituliskan dalam bahasa inggris, sehingga perlu diterjemahkan menjadi bahasa indonesia agar seragam. Proses penerjemahan data dilakukan dengan memanfaatkan library deep_translator dalam bahasa pemrograman Python.
- b. Case folding merupakan langkah pengubahan seluruh tanggapan ke dalam huruf kecil (lowercase) (Isnain et al., 2022). Pada proses ini karakter-karakter 'A'-'Z' yang terdapat pada data diubah menjadi karakter 'a'-'z'. Sementara itu, karakter lain yang tidak termasuk huruf dan angka, seperti tanda baca dan spasi, dianggap sebagai pemisah.
- c. Tokenizing merupakan langkah penghapusan data yang mengandung tanda baca, sehingga menghasilkan kalimat/kata yang berdiri sendiri (Isnain et al., 2022). Token dapat merujuk pada berbagai entitas, seperti kata, angka, simbol, tanda baca, dan lain-lain. Artinya, tahap ini bertujuan memecah tanggapan menjadi satuan kata.
- d. Stopword merupakan tahap dalam pengolahan teks yang melibatkan penghapusan kata-kata umum yang sering muncul dalam jumlah banyak dan tidak memiliki arti penting. Proses ini dapat dilakukan menggunakan algoritma stoplist untuk menghilangkan kata-kata yang dianggap kurang penting, atau dengan menggunakan wordlist untuk menyimpan daftar kata-kata yang dianggap penting (Isnain et al., 2022). Contohnya, kata-kata penghubung seperti 'dan', 'yang', 'serta', 'setelah', dan sejenisnya. Penghapusan stopwords ini memiliki manfaat seperti mengurangi ukuran indeks dan waktu pemrosesan teks, serta dapat mengurangi tingkat noise dalam data teks.
- e. Stemming merupakan langkah pengubahan setiap kata imbuhan menjadi kata dasar (Isnain et al., 2022). Tahap ini penting untuk mengurangi jumlah indeks

yang berbeda dari satu data dengan mengonversi kata-kata yang memiliki suffix maupun prefix kembali ke bentuk dasarnya. Selain itu, proses ini juga membantu dalam mengelompokkan kata-kata lain yang memiliki kata dasar dan arti yang serupa, namun memiliki bentuk yang berbeda karena menerima imbuhan yang berbeda pula.

2. Pembobotan Term Frequency - Inverse Document Frequency

Term Frequency-Inverse Document Frequency (TF-IDF) merupakan teknik pemberian bobot hubungan suatu term terhadap sebuah dokumen (Alfarizi et al., 2022). Metode TF-IDF mengkombinasikan dua konsep untuk menghitung bobot suatu term, yaitu frekuensi kemunculan sebuah term dalam sebuah dokumen (TF) dan inversi frekuensi dokumen (IDF) yang mengandung kata tersebut (Humaini et al., 2020).

Frekuensi kemunculan kata dalam sebuah dokumen menunjukkan tingkat pentingnya kata tersebut di dalam dokumen tersebut. Sementara itu, jumlah frekuensi dokumen yang mengandung kata tersebut menunjukkan seberapa umum kata tersebut secara keseluruhan. Semakin jarang kata tersebut muncul di dokumen-dokumen lain, maka nilai IDF-nya akan semakin tinggi.

3. Visualisasi Data Menggunakan Vector Space Model

Pada tahap ini, data akan ditampilkan menggunakan aplikasi web. Selain menampilkan hasil perangkingan dokumen dan query, tahap ini juga akan menampilkan seluruh hasil web scraping dari seluruh website yang terintegrasi dengan database. Vector Space Model (VSM) bertujuan menilai tingkat kedekatan atau kesamaan (similarity) antara term dengan menerapkan pembobotan pada setiap term (Pratama et al., 2020). Dalam metode ini, dokumen dianggap sebagai vektor-vektor yang memiliki jarak (magnitude) dan arah (direction).

Setiap term direpresentasikan sebagai dimensi dalam ruang vektor (Sidorov & Sidorov, 2019). Biasanya, term yang digunakan berdasarkan pada query atau keyword. Relevansi dokumen terhadap sebuah query diukur berdasarkan kesamaan antara vektor dokumen dan vektor query.

Pembobotan TF-IDF dan VSM digunakan untuk mewakili nilai numerik dokumen sehingga memungkinkan perhitungan

kedekatan antara dokumen-dokumen (Zen et al., 2021). Semakin dekat dua vektor dalam VSM, maka semakin mirip dua dokumen yang diwakili oleh vektor tersebut. Untuk menghitung kemiripan antara dokumen, digunakan suatu fungsi ukuran kemiripan (similarity measure) (Azis et al., 2019). Ukuran ini memungkinkan perangkingan dokumen berdasarkan tingkat kemiripan relevansinya terhadap query.

4. Evaluasi Recall dan Precision

Berdasarkan penelitian yang telah dilakukan sebelumnya, pengujian vector space model akan menggunakan evaluasi recall dan precision. Precision dapat dianggap sebagai ukuran ketepatan atau ketelitian, sedangkan recall adalah kesempurnaan. Pada penelitian ini, nilai Recall diperoleh berdasarkan dokumen terpanggil yang sesuai dengan permintaan pengguna. Sedangkan, nilai Precision diperoleh berdasarkan jumlah dokumen yang terpanggil dari database relevan setelah dinilai pengguna dengan informasi yang dibutuhkan.

HASIL DAN PEMBAHASAN

A. Preprocessing Data dan Pembobotan TF-IDF

Dataset berisi teks yang diperoleh dari 3 website freelance, dengan menggunakan bahasa pemrograman Python. Untuk proses web scraping menggunakan BeautifulSoup dan Selenium. Proses web scraping pada Projects dan Freelancer menggunakan BeautifulSoup sedangkan untuk Sribulancer menggunakan BeautifulSoup dan Selenium, karena untuk website Sribulancer memerlukan proses login dan hampir keseluruhan datanya ditampilkan menggunakan javascript. Selanjutnya, dataset dilakukan pra-proses data yang mencakup translate, case folding, stopword, stemming, dan tokenizing. Berikut adalah contoh hasil akhir pra proses data yang ditunjukkan pada Tabel 1.

Tabel 1. Hasil pra-proses data

Data Sebelum		Data Sesudah	
perlu	dokumen	[perlu],	[dokumen],
terjemah	bahasa	[terjemah],	[bahasa],
spanyol	bahasa inggris	[spanyol], [bahasa], [inggris],	
butuh	sertifikasi	[butuh], [sertifikasi],	
sedang	cari terjemah	[sedang], [cari], [terjemah],	
dapat	terjemah bahasa	[dapat], [terjemah], [bahasa],	
spanyol	bahasa inggris	[spanyol], [bahasa], [inggris],	
kiriman	format pdf	[kiriman], [format], [pdf],	
kurang	halaman butuh	[kurang], [halaman], [butuh],	
tawar	freelancer baru	[tawar], [freelancer], [baru],	
untuk	proyek harga	[untuk], [proyek], [harga],	
negosiasi		[negosiasi],	

Sumber: (Nurkholis et al., 2023)

Setelah praproses data selesai, maka dilanjutkan dengan pembobotan TF-IDF untuk mengubah data teks menjadi data numerik supaya bisa dilakukan perhitungan jumlah kemunculan kata dan juga untuk menghitung bobot setiap kata. Berikut contoh hasil pembobotan TF-IDF yang dapat dilihat pada Gambar 3.

(0, 199)	0.04902380648572567
(0, 93)	0.10905413768023318
(0, 279)	0.10905413768023318
(0, 1339)	0.11583843857672849
(0, 920)	0.10379182507824834
(0, 698)	0.12540037364272685
(0, 320)	0.10379182507824834
(0, 289)	0.07951068765630527
(A, 12A)	A.10905413768023318

Sumber: (Nurkholis et al., 2023)

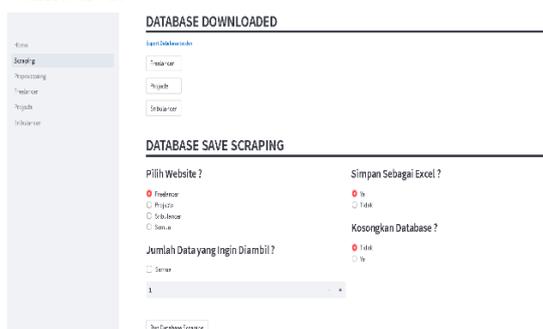
Gambar 3. Pembobotan TF-IDF

Nilai pertama pada Gambar 3 merepresentasikan indeks dari dokumen. Nilai kedua merepresentasikan indeks dari term (kata), sedangkan nilai ketiga merepresentasikan hasil perhitungan IDF (inverse document frequency).

B. Visualisasi Data Menggunakan Metode VSM

Pada tahap ini, dikembangkan user interface berbasis web. Tahap ini bertujuan menampilkan data hasil scraping dan mempermudah pengguna untuk menavigasi proses scraping data dan memproses data. Pengembangan user interface menggunakan library streamlit pada Python yang terdiri dari tiga menu utama, yakni pencarian data, preprocessing data, dan hasil scraping. Berikut adalah contoh halaman yang telah dikembangkan.

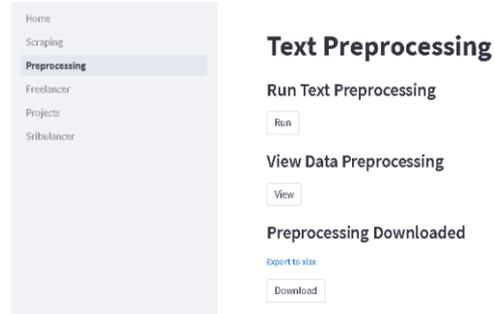
1. Halaman Scraping Data berfungsi melakukan proses web scraping dan mengunduh hasil impor scraping data. Pengguna dapat memilih sumber data yang akan di scraping, jumlah datanya, opsi untuk melakukan proses export data dari basis data ke dalam excel dan opsi untuk mengosongkan database saat melakukan scraping. Tampilan menu dapat dilihat pada Gambar 4.



Sumber: (Nurkholis et al., 2023)

Gambar 4. Menu pilihan web scraping

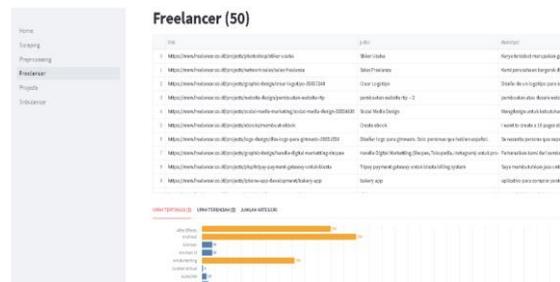
2. Halaman Preprocessing Data berfungsi melakukan pra proses data serta mengunduh hasilnya yang sudah di export kedalam excel. Data hasil web scraping harus melalui tahap pra proses ini untuk dapat melakukan pencarian data.



Sumber: (Nurkholis et al., 2023)

Gambar 5. Menu praproses data

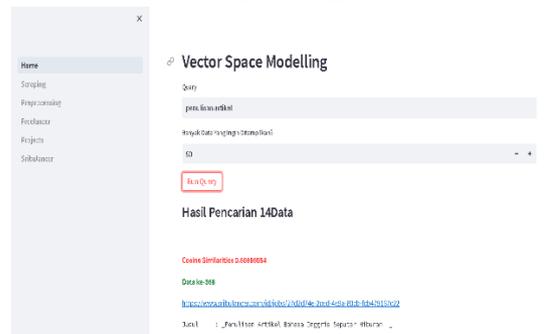
3. Tampilan Halaman Hasil Web Scraping pada Gambar 6 merupakan contoh hasil scraping dari situs Freelancer. Halaman ini juga memberikan informasi berupa grafik upah berdasarkan tiap kategori pekerjaan.



Sumber: (Nurkholis et al., 2023)

Gambar 6. Menu hasil web scraping

4. Halaman Pencarian Data berfungsi untuk menampilkan hasil pencarian freelance job dengan memasukkan query, ditunjukkan pada Gambar 7. Pengguna juga dapat memilih jumlah data yang ingin ditampilkan. Hasil pencarian data diurutkan berdasarkan nilai cosine similarities dari yang tertinggi ke terendah.



Sumber: (Nurkholis et al., 2023)

Gambar 7. Menu pencarian data VSM

C. Evaluasi Model Visualisasi

Pengujian dilakukan dengan menghitung nilai recall dan precision untuk mengevaluasi kinerja algoritma dalam mencari dokumen yang relevan. Pengujian ini melibatkan 31 dokumen dan menggunakan 5 kata kunci yang terkait dengan dokumen tersebut. Hasil dari pengujian, termasuk nilai recall dan precision ditunjukkan pada Tabel 2.

Tabel 2. Hasil evaluasi

Kata Kunci	Relevan	Tidak Relevan	Tidak Ditemukan	Recall (%)	Precision (%)
Program web	5	1	0	100	84
Desain grafis	11	6	0	100	65
Penulisan artikel	7	7	0	100	50
Edit video	5	5	0	98	50
Terjemah bahasa inggris	4	9	0	96	31
Rata-rata				98	56

Sumber: (Nurkholis et al., 2023)

Hasil pengujian pada Tabel 2 menunjukkan bahwa nilai rata-rata recall adalah 0,98 atau 98%, sementara nilai rata-rata precision adalah 0,56 atau 56%. Perbedaan ini terkait dengan jumlah dokumen yang berhasil diindeks. Semakin banyak dokumen yang berhasil diindeks, maka jumlah dokumen yang relevan juga meningkat, yang berdampak positif pada nilai recall. Namun, dengan peningkatan jumlah dokumen, juga meningkatkan kemungkinan munculnya noise (dokumen yang tidak relevan), sehingga tingkat precision menjadi rendah.

Selain itu dikarenakan penulis menggunakan 3 parameter untuk pengambilan data berdasarkan judul, deskripsi dan kategori sangat mempengaruhi nilai recall dan precision. Sehingga kemungkinan untuk menemukan data tentu saja akan semakin besar karena menggunakan 3 parameter, namun hal ini sangat mempengaruhi dalam tingkat precision karena jika didalam dokumen yang tidak relevan mengandung sebuah data dalam query maka dokumen tersebut pasti akan terambil walaupun sebenarnya kata didalam dokumen tersebut tidak relevan dalam konteks yang dimaksudkan oleh pengguna.

Pada kasus pencarian data dengan kata kunci "Program web" berhasil menemukan semua dokumen yang relevan hal ini karena dari semua dokumen tersebut mengandung kata "program" atau "web" dari 3 parameter yang digunakan. Sedangkan untuk dokumen yang tidak relevan ditemukan sebuah dokumen dengan judul "Dicari Blog/web Indo ataupun bule untuk di beli". Dokumen tersebut dapat ditampilkan saat pencarian data karena dari judul dokumen tersebut mengandung kata "web" walaupun dokumen

tersebut berisi tentang pembelian website dan tidak sesuai dengan yang dimaksudkan oleh pengguna.

KESIMPULAN

Penelitian ini berhasil menerapkan metode web scraping untuk pengambilan data dari 3 freelance. Namun dalam melakukan proses web scraping pada situs Sribulancer memerlukan koneksi yang baik dikarenakan data ditampilkan secara dinamik dengan javascript. Dalam penampilan data menggunakan framework streamlit untuk penyajian data dan membantu pengguna untuk menavigasi proses pengambilan dan pemrosesan data. Proses pencarian data menggunakan metode Vector Space Model pada hasil scraping menghasilkan rata-rata nilai recall yang mendekati sempurna, yaitu 0,98 atau 98%, Sedangkan untuk rata-rata nilai precision sebesar 56%.

Dalam proses pengambilan data menggunakan 3 parameter yaitu: judul, deskripsi dan kategori data lowongan freelance untuk meningkatkan hasil recall dalam pencarian data, namun hal ini sangat mempengaruhi nilai precision karena kemungkinan untuk mengambil data yang tidak relevan juga semakin besar jika dokumen tersebut mengandung sebuah kata pada query user walaupun konteksnya tidak sesuai. Penelitian menghasilkan sebuah aplikasi web yang dapat melakukan proses web scraping, pencarian informasi lowongan kerja freelance dari beberapa website secara sekaligus dan menampilkan hasil pemrosesan dataset dalam bentuk yang lebih ringkas dan bermanfaat.

REFERENSI

Alfarizi, M. I., Syafaah, L., & Lestandy, M. (2022). Emotional Text Classification Using TF-IDF (Term Frequency-Inverse Document Frequency) And LSTM (Long Short-Term Memory). *JUITA: Jurnal Informatika*, 10(2), 225-232.

Anna, A., & Hendini, A. (2018). Implementasi vector space model pada sistem pencarian mesin karaoke. *Evolusi : Jurnal Sains Dan Manajemen*, 6(1). <https://doi.org/10.31294/evolusi.v6i1.3535>

Ayoobzadeh, M. (2022). Freelance job search during times of uncertainty: protean career orientation, career competencies and job search. *Personnel Review*, 51(1), 40-56.

Azis, M. A., Hamid, A., Fauzi, A., Yulianto, E., & Riyanto, V. (2019). Information retrieval system in text-based skripsi document search file using vector space model method. *Journal of Physics: Conference Series*, 1367(1), 012016.

- Belwal, R. C., Rai, S., & Gupta, A. (2021). Text summarization using topic-based vector space model and semantic measure. *Information Processing & Management*, 58(3), 102536.
- Eminagaoglu, M. (2022). A new similarity measure for vector space models in text classification and information retrieval. *Journal of Information Science*, 48(4), 463–476.
- Firrezqi, W. A. (2020). Peran Situs Freelance Project. co. id Dalam Membantu Masalah Perekonomian di Indonesia. *Analisis Peran Situs Freelance Project. Co. Id Dalam Membantu Masalah Perekonomian Di Indonesia*, 2(2), 1–8.
- Han, S., & Anderson, C. K. (2021). Web scraping for hospitality research: Overview, opportunities, and implications. *Cornell Hospitality Quarterly*, 62(1), 89–104.
- Henrys, K. (2021). Importance of web scraping in e-commerce and e-marketing. Available at SSRN 3769593.
- Humaini, I., Wulandari, L., Ikasari, D., & Yusnitasari, T. (2020). Penerapan Algoritma TF-IDF Vector Space Model (VSM) Pada Information Retrieval Terjemahan Al Quran Surat 1 Samai Dengan Surat 16 Berdasarkan Kesamaan Makna. *Prosiding-Seminar Nasional Teknik Elektro UIN Sunan Gunung Djati Bandung*, 525–534.
- Isnain, A. R., Sulistiani, H., Hurohman, B. M., Nurkholis, A., & Styawati, S. (2022). Analisis Perbandingan Algoritma LSTM dan Naive Bayes untuk Analisis Sentimen. *JEPIN (Jurnal Edukasi Dan Penelitian Informatika)*, 8(2), 299–303.
- Julianto, Y., Setiabudi, D. H., & Rostianingsih, S. (2022). Analisis Sentimen Ulasan Restoran Menggunakan Metode Support Vector Machine. *Jurnal Infra*, 10(1), 1–7.
- Kadam, S., Shinde, S., Sharma, A., Mali, S., & Student, B. E. (2018). Price comparison of computer parts using web scraping. *Int. J. Eng. Sci.*
- Khder, M. A. (2021). Web Scraping or Web Crawling: State of Art, Techniques, Approaches and Application. *International Journal of Advances in Soft Computing & Its Applications*, 13(3).
- Mustofa, M. (2018). Pekerja Lepas (Freelancer) dalam Dunia Bisnis. *Jurnal MoZaiK*, 10(1), 19–25.
- Nurkholis, A., Alita, D., & Munandar, A. (2022). Comparison of Kernel Support Vector Machine Multi-Class in PPKM Sentiment Analysis on Twitter. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 6(2).
- Pratama, S. E., Darmalaksana, W., Maylawati, D. S., Sugilar, H., Mantoro, T., & Ramdhani, M. A. (2020). Weighted inverse document frequency and vector space model for hadith search engine. *Indones. J. Electr. Eng. Comput. Sci*, 18(2), 1004–1014.
- Ridwan, R., & Hermawan, T. A. (2019). Penerapan mesin pencari informasi dengan menggunakan metode Vector Space Model. *JURNAL TEKNIK INFORMATIKA (JUTEKIN)*, 7(2).
- Semendawai, J. N., Febiola, I., Pamungkas, B., & Ruliansyah, M. D. (2021). Perancangan Aplikasi Otomatisasi Menggunakan Bahasa Pemrograman Python Pada Aktivitas Monitoring Pemakaian Data Harian Kartu Internet Of Things. *Jurnal Rekayasa Elektro Sriwijaya*, 3(1), 193–198.
- Sidorov, G., & Sidorov, G. (2019). Vector Space Model for Texts and the tf-idf Measure. *Syntactic N-Grams in Computational Linguistics*, 11–15.
- Singh, R., & Singh, S. (2021). Text similarity measures in news articles by vector space model using NLP. *Journal of The Institution of Engineers (India): Series B*, 102, 329–338.
- Styawati, S., Nurkholis, A., Abidin, Z., & Sulistiani, H. (2021). Optimasi Parameter Support Vector Machine Berbasis Algoritma Firefly Pada Data Opini Film. *Jurnal RESTI (Rekayasa Sistem Dan Teknologi Informasi)*, 5(5), 904–910.
- Thomas, D. M., & Mathur, S. (2019). Data analysis by web scraping using python. *2019 3rd International Conference on Electronics, Communication and Aerospace Technology (ICECA)*, 450–454.
- Zen, B. P., Susanto, I., & Finaliamartha, D. (2021). TF-IDF Method and Vector Space Model Regarding the Covid-19 Vaccine on Online News. *Sinkron: Jurnal Dan Penelitian Teknik Informatika*, 6(1), 69–79.