

ANALISIS SISTEM REKOMENDASI PADA JOB RECOMMENDATION BERDASARKAN PROFIL LINKEDIN MENGGUNAKAN *COSINE SIMILARITY*

Roni Habibi¹, Muhammad Dzihan Albanna²

Program Studi D4 Teknik Informatika

email: roni.habibi@poltekpos.ac.id¹, muhammaddzihan55@gmail.com²

Abstrak

Kemajuan teknologi melahirkan era otomatisasi yang sudah banyak memberikan perubahan di berbagai sektor dalam kehidupan manusia. Namun ada beberapa hal yang tidak dapat diproses oleh mesin secara langsung, karena mesin tidak bisa melakukan proses pengolahan data secara otomatis. Permasalahan ini juga yang terdapat dalam alur perekrutan karyawan dimana proses filtering calon karyawan akan banyak memakan waktu, selain itu tingkat kecocokan data calon karyawan dengan data pekerjaan yang akan dilamar tidak bisa diukur. Maka untuk mendukung hal tersebut dibutuhkan keterampilan data *science* yang dapat mengubah sebuah data menjadi informasi. Begitu juga dalam masalah sehari-hari seperti mencari pekerjaan atau mencari karyawan, data *science* dapat memberikan rekomendasi *content-based recommendation*. Dalam permasalahan yang diangkat dalam penelitian ini maka metode yang akan digunakan adalah *content-based filtering* yang dilakukan dengan menghitung kemiripan antara konten. Selain itu dalam penelitian ini akan menggunakan pendekatan *Cosine Similarity* untuk membandingkan kedekatan antar teks atau *text matching*. Dengan perpaduan tersebut akan dihasilkan sebuah sistem rekomendasi pekerjaan yang visualisasinya akan ditampilkan dalam web menggunakan flask. Dengan adanya sistem rekomendasi pekerjaan diharapkan memberikan rekomendasi pekerjaan yang sesuai bagi para pencari kerja dan juga memberikan calon karyawan yang sesuai bagi HRD.

Kata Kunci: Sistem Rekomendasi, *Content-Based Recommendation*, *Cosine Similarity*

Abstrak

Technology has given birth to an era of automation that has provided changes in various sectors of human life. However, there are some things that cannot be processed by the machine directly, because the machine cannot process data automatically. This problem is also contained in a hole that will be seen where the process of screening prospective employees will take a lot of time, besides the level of suitability of prospective employee data with the data for the job being applied for cannot be measured. So to support this, data science skills are needed that can turn data into information. Likewise in everyday problems such as finding a job or looking for employees, data science can provide content-based recommendations. In the problems raised in this research, the method that will be used is content-based filtering which is done by calculating the value between content. In addition, this study will use the Cosine Similarity approach to compare the closeness between texts or text matching. With this combination, a job recommendation system will be produced whose visualization will be displayed on the web using flask. With the job recommendation system, it is expected to provide suitable job recommendations for job seekers and also provide suitable candidates for HRD.

Keyword: Recommendation System, *Content-Based Recommendation*, *Cosine Similarity*

1. PENDAHULUAN [Times New Roman 11 tebal]

Dalam era sekarang ini muncul sebuah bidang profesi baru yaitu *data science*. *Data science* merupakan gabungan dari inferensi data, pengembangan algoritma dan juga

teknologi untuk memecahkan masalah analitik yang kompleks. Dalam *data science* juga terdapat analisis prediktif suatu data untuk difilter dan ditemukan data yang benar agar menghasilkan suatu data yang akurat sesuai dengan data yang sebenarnya. Ada banyak

algoritma yang terdapat dalam *data science* salah satunya adalah rekomendasi sistem[1].

Saat berselancar di platform digital seperti *facebook* seringkali bertemu dengan orang yang dikenal secara tidak sengaja, nyatanya hal itu bukanlah ketidaksengajaan, itu adalah fitur rekomendasi yang terdapat pada *facebook* agar dapat bertemu dengan orang yang dikenal yang menggunakan platform yang sama. Begitupun platform lain seperti *netflix*, *tokopedia*, *youtube* yang sama-sama menggunakan fitur rekomendasi yang muncul sesuai dengan apa yang sedang dicari dan dibutuhkan[2].

Sistem rekomendasi adalah sebuah pendekatan untuk masalah seperti memberikan hal-hal yang cocok untuk pelanggan meskipun mencari banyak item. Meskipun selera Orang bervariasi dari satu ke yang lain tetapi mereka juga mengikuti beberapa pola. Sistem rekomendasi dapat menyaring konten halaman web secara otomatis, mengkategorikan pesan newsgroup untuk selanjutnya memberikan rekomendasi. Sistem ini dapat melakukan analisis yang berhubungan dengan data produk melalui interaksi yang diberikan *user* terhadap produk tersebut untuk dihubungkan. Misalnya di *tokopedia* terdapat fitur like dan masukan ke keranjang, item yang disukai dan dimasukan ke dalam keranjang adalah handphone, maka saat kembali ke aplikasi *tokopedia* sistem rekomendasi akan memberikan rekomendasi handphone yang mirip dan mungkin disukai[3]. Dalam bidang pekerjaan keresahan saat mencari pekerjaan sering terjadi karena kurangnya kematangan karir, kematangan karir merupakan kesiapan individu dalam pemilihan karir, serta proses pengambilan keputusan karir yang sesuai dengan tugas perkembangan karir. Dalam permasalahan yang diangkat dalam penelitian sistem rekomendasi dibutuhkan sebagai sistem yang digunakan perusahaan untuk memberikan rekomendasi pekerjaan yang cocok bagi *employee* atau peserta *internship*[4].

Untuk mendukung hal tersebut terdapat proses filtering yang terbagi ke dalam 4 bagian yaitu *collaborative filtering*, *content-based filtering*, *demographic*, dan *hybrid*. Pada judul ini penulis menggunakan *content based filtering* yang tergantung pada pilihan pengguna sebelumnya. Deskripsi dan profil orientasi pengguna memainkan peran penting peran dalam penyaringan berbasis konten[5].

Pada judul ini penulis menggunakan bahasa pemrograman *python* dengan melibatkan beberapa *library NLP* seperti *pandas*, *CSV*, *word2vec*, *glove*. *GloVe* adalah suatu metode *unsupervised learning* dalam representasi istilah yg mengguguli contoh lainnya pada *word analogy*, *word similarity*, & *named entity recognitio*[6]. Setelah itu data akan diolah menggunakan metode *cosine similarity* yang biasa diterapkan secara luas di pencarian informasi dan studi terkait[7].

2. LANDASAN TEORI

Sistem rekomendasi bertujuan untuk memberikan sebuah rekomendasi yang memancing perhatian user misalnya rekomendasi barang, rekomendasi teman, rekomendasi film, musik dan lain-lain[7]. Sistem rekomendasi menggunakan data yang diolah menjadi sebuah informasi untuk membuat suatu rekomendasi bagi user. User memiliki profil yang menggambarkan pribadi secara jangka panjang kemudian *recommendation system* memberikan kepada user sebuah rekomendasi yang relevan berdasarkan kesamaan deskripsi, profil atau informasi lain yang mirip[8].

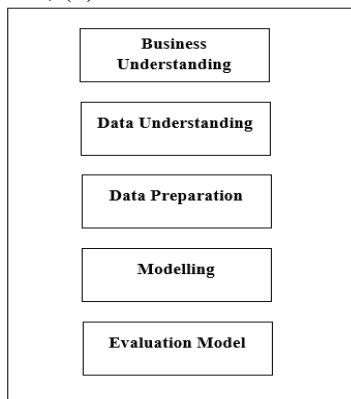
Pada algoritma ini menggunakan konten seperti atribut suatu item menjadi basis dalam pelabelan rekomendasi. Misalnya, dalam dataset *job* terdapat deskripsi dan *job title* yang memiliki konten atau isi teks didalamnya[9].

Algoritma *GloVe* juga menggabungkan informasi *co-occurrence* kata atau statistik global untuk memperoleh hubungan semantik antarkata dalam korpus. *GloVe* menggunakan metode *global matrix factorization*, matriks yang mewakili kemunculan atau ketiadaan kata-kata dalam suatu dokumen. Pelatihan model *GloVe* bertujuan untuk mempelajari vektor kata sedemikian rupa sehingga dot product katakata tersebut sama dengan logaritma probabilitas katakata untuk muncul bersama atau probabilitas *cooccurrence* nya[10]. *Cosine similarity* adalah suatu metode yang diterapkan secara luas di pencarian informasi dan studi terkait. Metode ini memodelkan teks sebagai vektor istilah dan kesamaan antara dua teks berasal nilai cosinus antara vektor istilah dua teks[7].

3. METODE PENELITIAN

Dalam penelitian ini juga mengadaptasi metode *CRISP-DM (CRoss-Industry Standard Process for Data Mining)*. Tahapan dari metode ini

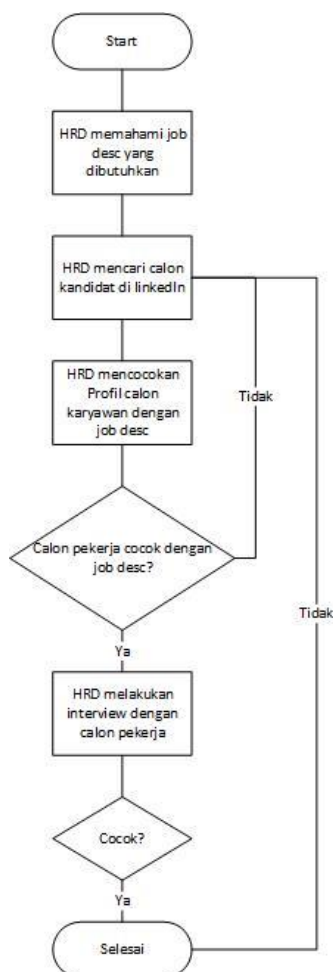
terdiri dari (1) Pemahaman Bisnis, (2) Pemahaman Data, (3) Persiapan Data, (4) Permodelan, (5) Evaluasi Model.



Gambar 1. Metode Penelitian

3.1 Business Understanding

Pada tahap ini observasi secara langsung dilakukan pada website perusahaan xyz. Kebutuhan yang diminta oleh perusahaan xyz yaitu fitur rekomendasi job yang akan diterapkan pada website. Alur yang saat ini berjalan di perusahaan xyz adalah sebagai berikut :



Gambar 2. Alur Bisnis Perekrutan Karyawan

Saat terdapat kebutuhan terhadap *job desc* tertentu HRD melakukan pencarian kandidat yang akan mengisi posisi tersebut. Namun sebelum itu HRD memahami terlebih dahulu *job desc* dari pekerjaan tersebut. Setelah itu HRD akan mencari kandidat karyawan di *LinkedIn* sebagai platform yang menyediakan berbagai macam profil pekerjaan secara profesional. Yang dilakukan pada saat pencarian kandidat ini adalah melihat profil calon karyawan seperti keahlian, pengalaman kerja, deskripsi, portofolio dan lain-lain. Jika dirasa cocok maka calon karyawan akan diarahkan untuk melakukan tes wawancara dengan HRD, apabila kriteria calon karyawan sesuai dengan *job desc* yang dibutuhkan maka tahapan pencarian karyawan selesai. Hal ini diangkat sebagai sesuatu yang dapat dipermudah dengan cara memberikan rekomendasi pekerjaan terhadap pihak-pihak terkait dengan menggunakan sistem rekomendasi, sehingga saat mencari karyawan yang cocok dengan *job desc* yang kosong berjalan lebih mudah. Untuk mendukung hal tersebut dikumpulkan data menggunakan *linkedhelper* dan dataset *job* sebagai pembandingan. Total jumlah keseluruhan data adalah 1638 data.

3.2 Data Understanding

Pemahaman mengenai data peneliti dapatkan berdasarkan kebutuhan dari metode-metode yang mendukung penelitian ini yang memerlukan *content* di dalam data yang akan digunakan dalam pembandingan. Maka peneliti menyiapkan data *job* dan *user*, kemudian dari kedua data tersebut terdapat kolom deskripsi yang dapat digunakan untuk *content based filtering*.

3.3 Data Preparation

Untuk memudahkan dalam tahap modelling maka dilakukan tahap data preparation yang membuat data menjadi dua kolom saja yang digunakan untuk pembandingan yaitu *about_clean* dan *desc_clean*. Sebelumnya data dibersihkan terlebih dahulu.

3.4 Modelling

Pemodelan dilakukan menggunakan *word embedding* dan *cosine similarity*. Model yang digunakan pada tahap *word embedding* adalah *Glove* yang dinilai lebih baik daripada *Word2vec* yang hanya mengandalkan informasi lokal dari kata-kata dengan konteks lokal. Hasil dari *glove* adalah berupa vektor yang berasal dari teks yang kemudian akan melalui tahap

similarity untuk memodelkan teks sebagai vektor istilah dan kesamaan antara dua teks berasal nilai *cosinus* antara vektor istilah dua teks

$$\cos(x, y) = x \cdot y / |x| * |y|$$

Contoh :

x mempunyai nilai $x = \{3,2,0,5\}$

y mempunyai nilai $y = \{1,0,0,0\}$

Rumus cosin similarity adalah :

$$\cos(x, y) = x \cdot y / |x| * |y|$$

$$x \cdot y = 3 * 1 + 2 * 0 + 0 * 0 + 5 * 0 = 3 \quad (1)$$

$$|x| = \sqrt{(3)^2 + (2)^2 + (0)^2 + (5)^2} = 6.16 \quad (2)$$

$$|y| = \sqrt{(1)^2 + (0)^2 + (0)^2 + (0)^2} = 1 \quad (3)$$

$$\cos(x, y) = 3 / (6.16 * 1) = 0.49 \quad (4)$$

3.5 Evaluation Model

Tahap terakhir yang dilakukan adalah pengujian yang dilakukan terhadap model untuk dievaluasi tingkat akurasi. Bagian ini mengukur bagaimana model tersebut menangani suatu masalah, memberikan solusi terhadap masalah dan tujuan yang ingin dicapai.

4. HASIL DAN PEMBAHASAN

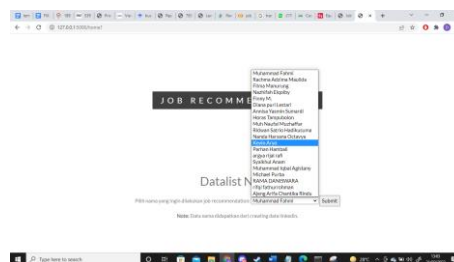
Berdasarkan proses algoritma yang telah dijelaskan diatas maka selanjutnya dilakukan pengujian terhadap yang sudah dibangun.



Gambar 3. Tampilan Home Job Recommendation

Data yang sudah dikumpulkan disimpan di *dropbox* sebagai alternatif untuk memudahkan user memilih data user mana yang akan digunakan untuk job rekomendasi.

Data yang terdapat pada *dropbox* hanya dataset *user* seperti yang tertera pada gambar 4.



Gambar 4. List Data User

Data dipilih kemudian disubmit untuk dicocokkan dengan dataset job. Hasil rekomendasi job akan muncul secara berurutan dari 5 data teratas dan 5 data terbawah seperti pada gambar 5.

Hasil Job Recommendation

Hasil 10 similarity teratas

	title	name	similarity
809	Data Scientist Machine Learning with Security Clearance	Nanda Harsana Octavyya	0.309305
837	Data Scientist - Cyber Incident Analysis (TS/SCI with Polygraph)	Nanda Harsana Octavyya	0.309305
821	Data Scientist (SME) - TS/SCI w/ Poly Required	Nanda Harsana Octavyya	0.309305
194	Machine Learning Scientist	Nanda Harsana Octavyya	0.309305
825	ORA/Data Scientist - AVAILABLE IMMEDIATELY	Nanda Harsana Octavyya	0.309305
192	Sr. Data Engineer, Marketing	Nanda Harsana Octavyya	0.309305
191	Senior Data Scientist, Machine Learning - Consumer & Merchant	Nanda Harsana Octavyya	0.309305
831	Biosurveillance Innovation Applied Scientist	Nanda Harsana Octavyya	0.309305
188	Assistant Research Scientist II (Pharmacology)	Nanda Harsana Octavyya	0.309305
197	Data Scientist Intern, Analytics (PhD)	Nanda Harsana Octavyya	0.309305

Hasil 10 similarity terbawah

	title	name	similarity
427	Marketing Analytics Manager @ FinTech Startup (\$B+ valuation)	Nanda Harsana Octavyya	0.293471
1609	Academic Data Analyst - Central Management Office	Nanda Harsana Octavyya	0.293260
1234	Big Data Engineer with very strong AWS	Nanda Harsana Octavyya	0.293260
178	Analytics Manager	Nanda Harsana Octavyya	0.293260
1215	Big Data Engineer (Training and Placement)	Nanda Harsana Octavyya	0.293260
548	Data Engineer lead / Jr. Architect	Nanda Harsana Octavyya	0.293260
15	RN/LPN (PRN)	Nanda Harsana Octavyya	0.293260
1564	Big Data Engineer with Cassandra and Spark	Nanda Harsana Octavyya	0.292991
27	NYS Licensed Psychologist	Nanda Harsana Octavyya	0.292699
1270	Data scientist	Nanda Harsana Octavyya	0.253056

Gambar 5. Hasil Rekomendasi Pekerjaan

Nilai dari similarity akan diurutkan atau dilakukan perankingan dari yang terbesar sampai terkecil. Adapaun patokan rekomendasi adalah rekomendasi yang terdapat di paling atas.

5. KESIMPULAN DAN SARAN

5.1 Simpulan

Berdasarkan hasil pengujian yang diperoleh dapat disimpulkan bahwa sistem rekomendasi menggunakan Metode *Cosine Similarity* dapat menghasilkan rekomendasi pekerjaan terhadap user yang diurutkan dari hasil similarity terbesar hingga terkecil sehingga memberikan kemudahan saat proses pencarian karyawan yang dibutuhkan untuk

mengisi *jobdesc* tertentu di perusahaan xyz. Dalam simulasi yang dilakukan dengan username A direkomendasikan beberapa pekerjaan yang dinilai sesuai diantaranya adalah *data scientist*, *machine learning scientist*, *senior data engineer* dan *marketing*.

5.2 Saran

Saran yang dibutuhkan oleh pengembang selanjutnya adalah akurasi dari sistem rekomendasi yang belum dapat diketahui. Pengembangan dari penelitian ini bisa dilakukan dengan menambahkan metode lain yang dapat memberikan nilai akurasi terhadap hasil rekomendasi pekerjaan.

6. DAFTAR PUSTAKA

- [1] F. Provost and T. Fawcett, *Data science and its relationship to big data and data-driven decision making*, vol. 1, no. 1. Mary Ann Liebert, Inc. 140 Huguenot Street, 3rd Floor New Rochelle, NY 10801 USA, 2013.
- [2] C. S. D. Prasetya, "Sistem Rekomendasi Pada E-Commerce Menggunakan K-Nearest Neighbor," *J. Teknol. Inf. dan Ilmu Komput. p-ISSN*, vol. 2355, p. 7699, 2017.
- [3] D. Das, L. Sahoo, and S. Datta, "A survey on recommendation system," *Int. J. Comput. Appl.*, vol. 160, no. 7, 2017.
- [4] A. Grashinta, A. P. Istiqomah, and E. P. Wiroko, "Pengaruh Future Time Perspective Terhadap Kematangan Karir Pada Mahasiswa," *J. Psikol. Pendidik. dan Konseling J. Kaji. Psikol. Pendidik. dan Bimbing. Konseling*, vol. 4, no. 1, p. 25, 2018, doi: 10.26858/jpkk.v4i1.4981.
- [5] A. E. Wijaya and D. Alfian, "Sistem Rekomendasi Laptop Menggunakan Collaborative Filtering Dan Content-Based Filtering," *J. Comput. Bisnis*, vol. 12, no. 1, pp. 11–27, 2018.
- [6] R. D. Indrapurasih, M. A. Bijaksana, I. L. Sardi, and L. Belakang, "Implementasi dan Analisis Kesamaan Semantik Antar Kata Bahasa Indonesia Menggunakan Metode GloVe," *eProceedings Eng.*, vol. 5, no. 3, pp. 7699–7706, 2018.
- [7] F. Rahutomo, T. Kitasuka, and M. Arisugi, "Semantic Cosine Similarity," *Semant. Sch.*, vol. 2, no. 4, pp. 4–5, 2012.
- [8] T. Badriyah, R. Fernando, and I. Syarif, "Sistem Rekomendasi Content Based Filtering Menggunakan Algoritma Apriori," *Konf. Nas. Sist. Inf.*, vol. 1, no. 1, pp. 554–559, 2018.
- [9] P. Nastiti, "Penerapan Metode Content Based Filtering Dalam Implementasi Sistem Rekomendasi Tanaman Pangan," *Teknika*, vol. 8, no. 1, pp. 1–10, 2019, doi: 10.34148/teknika.v8i1.139.
- [10] M. Ibrahim, S. Gauch, O. Salman, and M. Alqahatani, "Enriching consumer health vocabulary using enhanced GloVe word embedding," *arXiv Prepr. arXiv2004.00150*, 2020.