

Sistem Rekomendasi Content Based Filtering Pekerjaan dan Tenaga Kerja Potensial menggunakan Cosine Similarity

Philips Nogo Raharjo¹, Andreas Handojo², Hans Juwiantho³
Program Studi Informatika Fakultas Teknologi Industri Universitas Kristen Petra
Jl. Siwalankerto 121 – 131 Surabaya 60236
Telp. (031) – 2983455, Fax. (031) - 8417658

E-mail: philipsraharjo@gmail.com¹, handojo@petra.ac.id², hans.juwiantho@petra.ac.id³

ABSTRAK

Pada saat pandemi, terjadi masalah ekonomi yang membuat perusahaan harus melakukan sesuatu untuk menghindari kerugian yang ada. Salah satu usahanya adalah memutuskan hubungan kerja dengan tenaga kerjanya. Dengan cara tradisional, tenaga kerja dan perusahaan akan membuang banyak waktu untuk mencari yang cocok baginya. Sistem rekomendasi lowongan pekerjaan dan tenaga kerja potensial berperan penting di kondisi seperti ini. Karena dengan adanya sistem rekomendasi yang dapat membantu dari dua sisi akan mempercepat pertemuan antara perusahaan yang membutuhkan pekerja dan pekerja yang membutuhkan lowongan pekerjaan.

Berdasarkan pengujian yang telah dilakukan, sistem rekomendasi menggunakan model *TF-IDF* dapat memberikan rekomendasinya dengan baik berdasarkan perhitungan *Mean Reciprocal Rank* mendapatkan sebesar 0.857 dan *Mean Average Precision* sebesar 0.833, dimana hasil ini cukup baik.

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

ABSTRACT

During the pandemic, there was an economic problem that forced companies to do something to avoid any loss. One of the action is to terminate the employment with their workforces. In the conventional way, the workforce and the company will waste a lot of time looking for the right fit for them. So, the recommendation system for jobs and workforce plays an important role in these conditions. Because with the existence of recommendation system that can help from both sides, it will speed up the meeting between companies that need workers and workers who need jobs.

*Based on the test have been carried out, the recommendation system using the *TF-IDF* model can provide good recommendations based on the calculation of the *Mean Reciprocal Rank* getting 0.857 and *Mean Average Precision* of 0.833, where these results are quite good.*

Keywords: *Recommendation System, Content-Based Filtering, Cosine Similarity*

1. PENDAHULUAN

Pada saat ini, dunia sedang dilanda oleh virus COVID-19 yang menyebabkan munculnya permasalahan-permasalahan yang baru. Buruknya pertumbuhan ekonomi juga mempengaruhi situasi ketenagakerjaan di Indonesia. Aktivitas perekonomian yang terhambat menyebabkan para pelaku bisnis harus melakukan efisiensi untuk menekan kerugian yang timbul. Akibatnya, banyak

pekerja dan perusahaan yang dirugikan, baik karena kehilangan pekerjaan atau kehilangan pekerja. Sistem rekomendasi lowongan pekerjaan dan tenaga kerja potensial berperan penting di kondisi seperti ini. Karena dengan adanya sistem rekomendasi yang dapat membantu dari dua sisi akan mempercepat pertemuan antara perusahaan yang membutuhkan pekerja dan pekerja yang membutuhkan lowongan pekerjaan.

Berdasarkan data Kementerian Ketenagakerjaan (Kemnaker) per 7 April 2020, tercatat sebanyak 39.977 perusahaan di sektor formal yang merumahkan atau melakukan PHK terhadap pekerjanya, dengan total sebanyak 1.010.579 orang pekerja yang terkena dampak ini. Sedangkan perusahaan di sektor informal ada sebanyak 34.453 perusahaan dan 189.452 orang pekerja [6]. Tim Riset SMERU menyebut bahwa angka yang dicatat masih belum mencerminkan angka pengangguran yang sesungguhnya karena belum mencakup semua sektor dan angkatan baru yang masih menganggur. Oleh karena itu, muncul masalah baru di pasar tenaga kerja, dimana pertama penyerapan tenaga kerja yang terjadi tidak sebanding dengan jumlah tenaga kerja yang terkena PHK. Berikutnya, perusahaan akan mulai mengubah kebijakan-kebijakannya mengenai rekrutmen tenaga kerja di masa yang mendatang. Sebagai contoh, perusahaan akan cenderung memilih tenaga kerja yang memiliki tingkat produktivitas tinggi dan dapat melakukan *multitasking*. Ini merupakan salah satu contoh tindakan perusahaan untuk melakukan efisiensi dalam pengelolaan keuangan.

Pada penelitian sebelumnya dari Gupta & Garg, penulis bertujuan untuk mengaplikasikan dan membandingkan metode-metode *data mining* yang ada untuk membuat sistem rekomendasi pekerjaan kepada tenaga kerja, dengan menggunakan CBR (*Content Based Recommender*), Model Based CFR (*Collaborative Filtering Recommender*), dan KBR (*Knowledge Based Recommender*). Penelitian ini menggunakan input *user information* dan *behavior* hanya memberikan rekomendasi kepada sisi tenaga kerja [4]. Selanjutnya penelitian sebelumnya dari Valverde-Rebaza et al, penulis bertujuan untuk mengaplikasikan berbagai algoritma *text-processing* seperti variasi *Word2vec*, dan *TF-IDF* dalam merekomendasi suatu jenis pekerjaan kepada tenaga kerja [8]. Untuk penelitian sebelumnya dari Malinowski et.al, penulis bertujuan untuk membangun suatu pendekatan dengan menggunakan dua sistem rekomendasi berbeda yang akan diharapkan dapat menambah tingkat akurasi antara pekerja dan lowongan pekerjaan. Model yang dipakai adalah *probabilistic hybrid recommendation* [5].

Oleh karena itu untuk mengatasi masalah lamanya proses dalam pencarian lowongan pekerjaan maupun tenaga kerja potensial, pada

skripsi ini mengusulkan untuk membuat sistem rekomendasi yang dapat mencocokkan preferensi tenaga kerja dengan tawaran kerja, dan mencari tenaga kerja potensial yang sesuai dengan lowongan pekerjaan. Skripsi ini akan menggunakan *Content-Based Filtering* dengan algoritma TF-IDF dan *Word2Vec-CBOW*, digunakan *Content-Based Filtering* karena tidak membutuhkan data dari user lain, sehingga rekomendasi dibuat spesifik untuk tiap user dan ini memudahkan apabila jumlah user sangat banyak. *Cosine Similarity* juga digunakan untuk menghitung sudut yang terbentuk. Kelebihannya meskipun ada data yang mirip namun memiliki jarak yang jauh, maka sudut yang terbentuk akan kecil dan perhitungannya jauh lebih baik daripada menggunakan metode *Euclidean Distance* dan lainnya. TF-IDF digunakan karena salah satu metode *text-processing* yang simpel dan relevan. Lalu *Word2Vec-CBOW* juga digunakan karena merupakan salah satu algoritma *word embedding* yang dapat mengetahui hubungan semantik antar kata.

2. PENELITIAN SEBELUMNYA

Penelitian akan menggunakan beberapa penelitian lain terkait yang telah dilakukan sebelumnya sebagai tinjauan studi. Berikut adalah penelitian-penelitian yang telah dilakukan dalam sistem rekomendasi:

2.1 Applying Data Mining Techniques in Job Recommender System for Considering Candidate Job Preferences

Gupta & Garg [4] melakukan penelitian yang bertujuan untuk mengaplikasikan metode *data mining* yang dapat memberikan list rekomendasi pekerjaan yang cocok dengan pengguna. Metode yang diusulkan dari penelitian ini adalah menggunakan CBR(*Content Based Recommender*), dan Model Based CFR(*Collaborative Filtering Recommender*), KBR(*Knowledge Based Recommender*). Hasil dari penelitian ini adalah menunjukkan bahwa penggunaan KBR(*Knowledge Based Recommender*) dalam pembuatan sistem rekomendasi pekerjaan memiliki nilai akurasi yang tinggi. Penggunaan metode tersebut dirasa sangat bagus, namun memiliki kekurangan yaitu masalah *data sparsity* pada saat awal.

2.2 Job Recommendation based on Job Seeker Skills: An Empirical Study

Valverde-Rebaza et.al [8] melakukan penelitian yang bertujuan untuk memberikan rekomendasi lowongan pekerjaan kepada *job seeker* berdasarkan skill yang dimiliki. Metode yang digunakan adalah berbagai metode *text processing*, seperti : TF-IDF dan 4 variasi *Word2vec*. Hasil dari penelitian ini menunjukkan bahwa hasil tertinggi didapatkan *Word2vec-SkipGram* dengan skor 0.590 dan disusul oleh TF-IDF pada peringkat dua dengan skor 0.588.

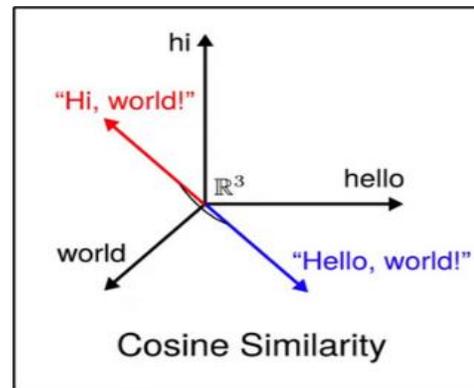
2.3 Matching People and Jobs: A Bilateral Recommendation Approach

Malinowski et.al [5] melakukan penelitian yang bertujuan untuk membangun suatu pendekatan dengan menggunakan dua sistem rekomendasi berbeda yang akan diharapkan dapat menambah tingkat akurasi antara pekerja dan lowongan pekerjaan. Metode yang diusulkan dari penelitian ini adalah menggunakan *Probabilistic Hybrid Recommendation*. Hasil dari penelitian ini adalah menyebutkan implementasi sistem berjalan dengan baik dilihat dari *feedback* yang didapat dari survei.

2.4 Cosine Similarity

Metode yang digunakan dalam penelitian ini terbagi menjadi dua model, yaitu TF-IDF dan *Word2Vec*. Dimana tiap-tiap model akan

dihitung menggunakan *cosine similarity*. *Cosine Similarity* adalah metode untuk menghitung kemiripan antara dua vektor dengan mencari *cosine* dari sudut tersebut. *Cosine Similarity* memiliki keuntungan yaitu meskipun dua dokumen yang mirip berjarak jauh oleh metode lain seperti *Euclidean Distance*, namun bisa dianggap masih dekat.



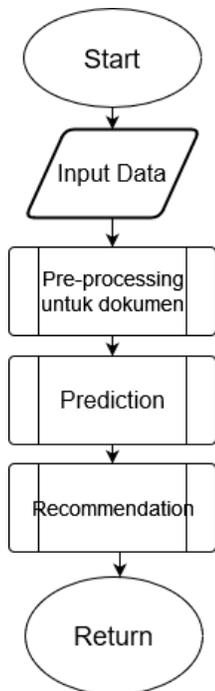
Gambar 1. Visualisasi Cosine Similarity

Pada Gambar 1 ditunjukkan visualisasi dari *Cosine Similarity*. Semakin kecil derajat yang dibentuk, maka semakin besar nilai dari *cosine similarity* [3]. Berikut rumus yang digunakan untuk mencari sudut kemiripan yang tertera pada Rumus 1 :

$$\cos \theta = \frac{\vec{a} \cdot \vec{b}}{\|\vec{a}\| \|\vec{b}\|} = \frac{\sum_1^n a_i b_i}{\sqrt{\sum_1^n a_i^2} \sqrt{\sum_1^n b_i^2}} \quad (1)$$

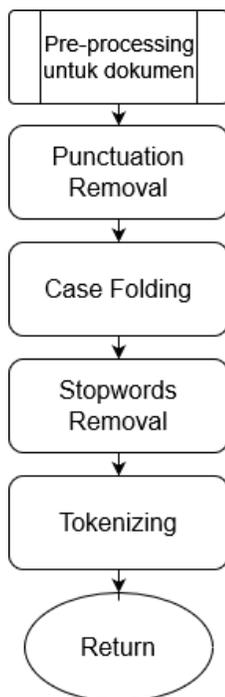
Content Based Filtering digunakan pada penelitian ini karena cocok dengan dataset yang dimiliki. *Collaborative Based Filtering* kurang tepat digunakan karena membutuhkan data perilaku dari user lain. Sedangkan sistem yang dibuat masih baru, yang artinya belum memiliki informasi yang cukup, sehingga mengakibatkan hasil yang tidak sesuai. Masalah tersebut disebut sebagai *cold start problem*, yang dapat diminimalisir dengan penggunaan *Content Based Filtering* [2]. Data informasi preferensi pengguna harus dipastikan adalah data yang baik dan relevan. Berdasarkan algoritma tersebut, prediksi akan dibuat berdasarkan item yang saling terkait ataupun memiliki tingkat kemiripan yang tinggi. Dalam kasus ini, data yang dipakai adalah *resume job seeker* dan dokumen detail lowongan pekerjaan *job provider*. Keunggulan dari *content based recommender* adalah tidak memerlukan data dari pengguna lain, karena model dibuat sesuai masing-masing preferensi pengguna. Sedangkan kelemahannya adalah model tetap memberikan rekomendasi berdasarkan kesukaan pengguna yang akan menimbulkan potensi pembatasan pengguna untuk mengeksplor hal baru yang mungkin disukainya [9].

Dalam *flowchart* pada Gambar 2 akan dijelaskan mulai dari proses awal *input*, proses, hingga *output* yang dihasilkan. Secara umum berikut *flowchart* dari sistem rekomendasi lowongan pekerjaan dan tenaga kerja potensial yang dibuat.



Gambar 2. Flowchart Rekomendasi

Pada awalnya, user akan mengupload file antara *resume* atau *job posting* sesuai dengan *role* yang dipilihnya. File yang diupload berbentuk *pdf*. Setelah file diupload, maka file tersebut akan masuk ke dalam tahap pre-processing data. Dimana pada proses ini, akan dilakukan proses *punctuation removal* yaitu proses penghilangan tanda baca.



Gambar 3. Flowchart Preprocessing

Sesuai dengan Gambar 3, proses pre-processing data dilakukan dengan tahap pertama yaitu *punctuation removal*, yaitu proses menghilangkan tanda baca pada suatu teks. Selanjutnya *case folding*, yaitu membesarkan atau mengecilkan suatu teks agar menjadi satu tipe dan dapat diproses lebih baik. Proses *stemming* dilakukan untuk menemukan kata dasar dari suatu kata. Namun, pada penelitian ini, tidak dilakukan *stemming* dikarenakan *stemming* yang dilakukan menjadi *overstemming* atau malah menghilangkan makna asli dari suatu kata. *Stopwords Removal* yaitu proses menghilangkan kata-kata yang sangat umum dipakai, contohnya seperti kata penghubung atau kata ganti. Terakhir dilakukan *tokenizing* yaitu proses pemecahan teks menjadi vektor yang berisikan kata-kata. Dimana nantinya vektor akan menjadi input dalam model sistem rekomendasi.

3. DATASET

Dataset yang digunakan dalam penelitian ini merupakan dataset yang diambil dari website *Kaggle*, dimana terdapat dua jenis dataset: dataset *resume* dari tenaga kerja dan dataset *job posting*.

3.1 Dataset Resume dari Tenaga Kerja

Dataset berisi tentang *resume* tenaga kerja berbentuk PDF dengan jumlah 2484 file. Dataset memiliki isi sebagai berikut : posisi pekerjaan, *highlights*, *accomplishments*, *experience*, *education*, dan *skills*. Data *resume* yang ada cukup panjang, lengkap, dan detail mengenai pengalaman dan kemampuan apa yang dimiliki oleh tenaga kerja.

3.2 Dataset Job Posting

Dataset berisi tentang lowongan pekerjaan yang diambil dari website Yahoo! Mailing Group dan memiliki 18892 baris dan 24 kolom. Untuk sistem rekomendasi, kolom yang diambil adalah *jobpost*. Adapun isi dari kolom *jobpost* adalah nama perusahaan, *job title*, lokasi, *job description*, *job responsibilities*, *required qualifications*. Bentuk file yang ada di dataset berbentuk *csv*, yang nantinya akan diubah menjadi *pdf*.

3.3 TF-IDF

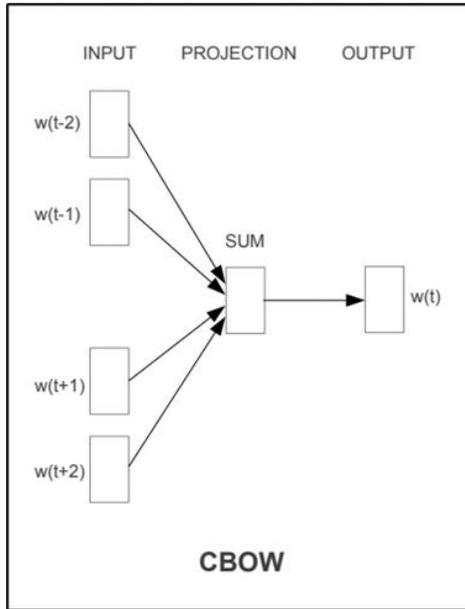
TF-IDF (*Term Frequency – Inverse Document Frequency*) adalah teknik pemberian bobot untuk setiap kata yang muncul untuk menandakan pentingnya kata tersebut di sebuah dokumen. Teknik ini sering digunakan pada *Text Mining* dan *Information Retrieval* [7]. TF mengukur frekuensi kata muncul dalam suatu dokumen. Namun hal tersebut tetap bergantung kepada panjangnya suatu dokumen. Contoh, kata hubung “dan” muncul di dua dokumen yang satu memiliki 100 kata, dan kedua memiliki 10.000 kata. Kata “dan” tidak bisa disebut lebih penting pada dokumen kedua karena jumlah frekuensinya yang lebih banyak. IDF adalah invers dari DF (*Document Frequency*) yang menunjukkan keamatan sebuah kata dengan dokumen. Dalam menghitung IDF, maka nilainya akan rendah untuk kata yang sering muncul seperti kata hubung. Berikut rumus dasar TF-IDF yang tertera pada Rumus 2 :

$$tf - idf(t, d) = tf(t, d) * \log(N / (df + 1)) \quad (2)$$

3.4 Word2Vec

Word2Vec adalah salah satu teknik *machine learning* yang menggunakan *neural network model* untuk mempelajari asosiasi kata dari suatu dokumen/*corpus*. Pada Gambar 4, ditunjukkan model arsitektur dari CBOW. *Word2Vec* adalah 2 layer *neural network* yang memproses teks dengan cara membuatnya menjadi

vektor. Inputnya adalah kata-kata dalam suatu dokumen dan outputnya adalah set berupa vektor. *Word2Vec* akan merepresentasikan tiap kata dengan angka tertentu yang disebut vektor. Setelah itu vektor akan dipilih untuk dihitung menggunakan *Cosine Similarity*. *CBOW* sendiri adalah model yang mempelajari dengan cara memprediksi kata dari suatu list konteks kata-kata yang ada [1].



Gambar 4. Model Arsitektur CBOW

4. Pengujian

Pengujian terhadap model dan sistem yang telah dibuat. Pengujian model dilakukan dengan menggunakan data 50 responden job seeker dan 50 responden job provider. Pengujian diterapkan pada 2 model, yaitu TF-IDF dan Word2Vec. Untuk proses pengujian rekomendasi, hasil rekomendasi yang diberikan berjumlah 10 rekomendasi, yaitu dimana 5 rekomendasi model TF-IDF dan 5 rekomendasi model Word2Vec ditampilkan di kolom berbeda.

Hasil dari rekomendasi yang diberi label oleh responden nantinya akan dihitung menggunakan 2 metode yaitu Mean Average Precision, dan Mean Reciprocal Rank. Pada pengujian model, akan dibagi menjadi 2 secara keseluruhan, yaitu pengujian model TF-IDF dan pengujian model Word2Vec. Masing-masing model akan dievaluasi menggunakan 2 metode, yaitu Mean Reciprocal Rank(MRR) dan Mean Average Precision(MAP)[10]. Tiap model juga dievaluasi dari dua sisi, yaitu job seeker dan job provider.

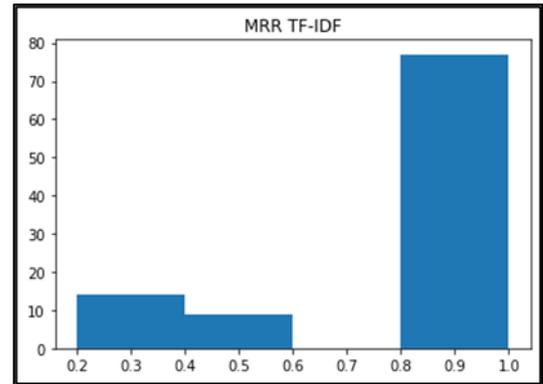
4.1 Pengujian Sistem

Sistem dibuat dalam bentuk *website* dengan bantuan framework *Flask*. Seperti layaknya sebuah *website*, *user* harus melakukan *sign up* dan *login*. Setelah itu *user* dapat melakukan upload file dalam bentuk *pdf*, dan mendapatkan hasil rekomendasi yang sesuai dengan preferensinya.

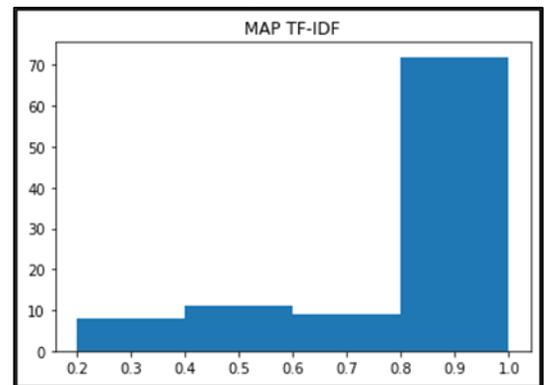
4.2 Pengujian model TF-IDF

Pengujian ini bertujuan untuk mencari seberapa efektif sistem rekomendasi yang menggunakan model *TF-IDF*. 50 responden yang sebagai *job seeker* akan menilai 5 hasil rekomendasi lowongan pekerjaan yang dihasilkan dari model *TF-IDF*. 50 responden yang sebagai *job seeker* juga akan menilai rekomendasi tenaga kerja potensial yang ditampilkan. Responden akan

memberikan penilaian “Yes” atau “No” terhadap masing-masing item rekomendasi. Berikut hasil penghitungan yang diperoleh. Berikut merupakan histogram MRR dan MAP dari TF-IDF pada Gambar 5 dan Gambar 6.



Gambar 5. Histogram MRR dari TF-IDF



Gambar 6. Histogram MAP dari TF-IDF

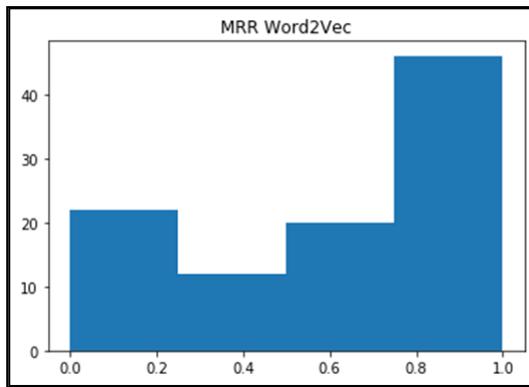
Dapat dilihat pada Tabel 1, ternyata secara pengujian model, nilai yang didapat dari *job seeker* lebih baik dibandingkan *job provider*. Ini bisa disebabkan oleh beberapa faktor, salah satunya adalah ketidakseimbangan data yang ada. *Job posting* yang ada di dataset mencapai 18992 baris, sedangkan *resume* yang ada di dataset hanya ada sekitar 2484 baris.

Tabel 1. Tabel Pengujian Model TF-IDF

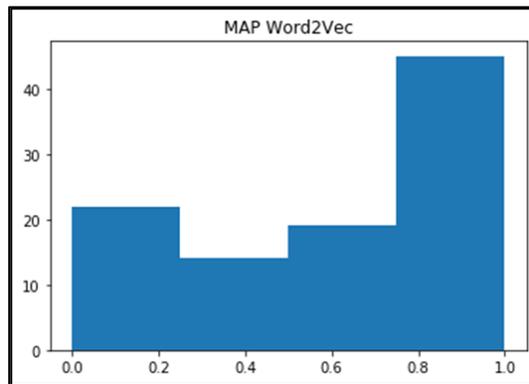
	<i>Job Seeker</i>	<i>Job Provider</i>
<i>Mean Reciprocal Rank</i>	0.8573	0.8567
<i>Mean Average Precision</i>	0.8495	0.8171

4.3 Pengujian model Word2Vec

Untuk pengujian ini, akan dilakukan pengujian terhadap model *Word2Vec CBOW*. Sama seperti dengan pengujian sebelumnya, akan dilakukan perhitungan *Mean Reciprocal Rank* dan *Mean Average Precision*. Berikut merupakan histogram MRR dan MAP dari *Word2Vec* pada Gambar dan Gambar 9.



Gambar 7. Histogram MRR dari Word2Vec



Gambar 8. Histogram MAP dari Word2Vec

Dapat dilihat pada Tabel 2, sama seperti sebelumnya, nilai pengujian dari sisi *job seeker* lebih baik. Namun, nilai yang didapat model *Word2Vec* dibandingkan dengan model *TF-IDF* terpaut cukup jauh.

Tabel 2. Tabel Pengujian Model *Word2Vec*

	<i>Job Seeker</i>	<i>Job Provider</i>
Mean Reciprocal Rank	0.6193	0.5963
Mean Average Precision	0.601583	0.56283

4.4 Hasil Pengujian Kedua Model

Untuk meringkas seluruh pengujian terhadap 2 model tersebut dan memahami lebih baik lagi mengenai kinerja kedua model sistem rekomendasi, maka akan disajikan Tabel 3 yang secara langsung membandingkan hasil perhitungan dari kedua model, yaitu *TF-IDF* dan *Word2Vec*.

Tabel 3. Tabel Perbandingan Pengujian Model *TF-IDF* dan *Word2Vec*

	<i>TF-IDF</i>	<i>Word2Vec</i>
Mean Reciprocal Rank	0.857	0.6078
Mean Average Precision	0.833	0.5822

Pada tabel di atas terlihat, bahwa model *TF-IDF* mengungguli model *Word2Vec* dengan cukup jauh. Untuk Model *TF-IDF*, *TF-*

IDF mendapatkan *Mean Reciprocal Rank* sebesar 0.857 dan *Mean Average Precision* sebesar 0.833. Sedangkan untuk model *Word2Vec*, hanya mendapatkan *Mean Reciprocal Rank* sebesar 0.6078 dan *Mean Average Precision* 0.5822.

Ada beberapa faktor yang menyebabkan rendahnya akurasi dari model *Word2Vec*. Pertama, yaitu *Word2Vec* tidak dapat menangani kata-kata yang keluar dari suatu *vocabulary*-nya. Oleh karena data yang dipakai adalah meliputi dunia kerja, di mana banyak sekali istilah yang tidak dipakai secara umum, semisal nama skill, benda, ataupun keterangan lain. Kedua, *Word2Vec* membutuhkan informasi yang banyak mengenai suatu kosakata, dikarenakan cara kerjanya yang memprediksi arti dan konteks dari suatu kata menurut kata-kata yang di sekitarnya.

Untuk penggunaan model *TF-IDF* pada penelitian ini tergolong cukup optimal, dikarenakan caranya yang simpel, mudah, dan juga menghitung frekuensi dari suatu kata itu muncul di dalam suatu dokumen untuk menentukan seberapa relevan kata tersebut. *TF-IDF* tidak memperdulikan urutan kata yang ada, oleh karena itu *TF-IDF* tidak dapat memahami konteks dari suatu kalimat.

5. KESIMPULAN

Pada penelitian ini, dilakukan pemodelan sistem rekomendasi *content-based filtering* pekerjaan dan tenaga kerja potensial menggunakan *cosine similarity*. Dimana terdapat dua model yang dikembangkan untuk bisa menghasilkan list rekomendasi, yaitu model *TF-IDF* dan model *Word2Vec*. Untuk dataset yang memiliki karakteristik seperti ini, model *TF-IDF* lebih diuntungkan karena tidak bergantung pada suatu *dictionaries* model yang ada di *Word2Vec*. Terbukti model *TF-IDF* memperoleh hasil *Mean Reciprocal Rank* 0.857, dan *Mean Average Precision* 0.833. Hal yang perlu diperhatikan juga adalah pada saat proses *pre-processing* data yang harus dilakukan lebih baik, dengan adanya format-format tertentu seperti email, nomor telepon, emoji, dan lainnya. Terakhir yang harus diperhatikan adalah keseimbangan data yang ada antara *job posting* dan juga *resume*. Keseimbangan tidak merujuk kepada hanya jumlah data saja, namun persebaran kategori data yang merata.

Berikut adalah beberapa saran yang dapat dilakukan untuk mengembangkan penelitian lebih lanjut:

1. Menambahkan metode Named-Entity Recognition dalam mengekstrak informasi yang ada dalam suatu dokumen, agar proses yang dilakukan dapat lebih ringan.
2. Pengolahan data teks yang lebih baik lagi untuk menghindari *noise* pada data.
3. Menggabungkan *collaborative filtering* dengan *content-based filtering* agar hasil rekomendasi dapat lebih beragam.

6. REFERENSI

- [1] Cahyani, D. E., & Patasik, I. (2021). Performance comparison of TF-IDF and Word2Vec models for emotion text classification. *Bulletin of Electrical Engineering and Informatics*, 10(5), 2780-2788. DOI: <https://doi.org/10.11591/eei.v10i5.3157>
- [2] Geetha, G., Safa, M., Fancy, C., & Saranya, D. (2018, April). A hybrid approach using collaborative filtering and content based filtering for recommender system. In *Journal of Physics: Conference Series* (Vol. 1000, No. 1, p. 012101). IOP Publishing. DOI:10.1088/1742-6596/1000/1/012101
- [3] Gunawan, D., Sembiring, C. A., & Budiman, M. A. (2018, March). The implementation of cosine similarity to calculate

- text relevance between two documents. In *Journal of physics: conference series* (Vol. 978, No. 1, p. 012120). IOP Publishing. DOI:10.1088/1742-6596/978/1/012120
- [4] Gupta, A., & Garg, D. (2014). Applying data mining techniques in job recommender system for considering candidate job preferences. *2014 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*. Published. <https://doi.org/10.1109/icacci.2014.6968361>
- [5] Malinowski, J., Keim, T., Wendt, O., & Weitzel, T. (2006). Matching People and Jobs: A Bilateral Recommendation Approach. *Proceedings of the 39th Annual Hawaii International Conference on System Sciences (HICSS'06)*. Published. <https://doi.org/10.1109/hicss.2006.266>
- [6] Rizal, J. G. (2020, August 11). *Pandemi Covid-19, Apa Saja Dampak pada Sektor Ketenagakerjaan Indonesia? Halaman all* - Kompas.com. KOMPAS.Com. <https://www.kompas.com/tren/read/2020/08/11/102500165/pandemi-covid-19-apa-saja-dampak-pada-sektor-ketenagakerjaan-indonesia-?page=all>
- [7] Qaiser, S., & Ali, R. (2018). Text mining: use of TF-IDF to examine the relevance of words to documents. *International Journal of Computer Applications*, 181(1), 25-29. DOI:10.5120/ijca2018917395
- [8] Valverde-Rebaza, J. C., Puma, R., Bustios, P., & Silva, N. C. (2018). Job Recommendation Based on Job Seeker Skills: An Empirical Study. In *Text2Story@ ECIR* (pp. 47-51).
- [9] Van Meteren, R., & Van Someren, M. (2000, May). Using content-based filtering for recommendation. In *Proceedings of the machine learning in the new information age: MLnet/ECML2000 workshop* (Vol. 30, pp. 47-56).
- [10] Zhang, L., Zhang, S., & Balog, K. (2019, July). Table2vec: Neural word and entity embeddings for table population and retrieval. In *Proceedings of the 42nd International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 1029-1032). DOI: <https://doi.org/10.48550/arXiv.1906.00041>