

IMPELEMNTASI NAÏVE BAYES DAN COSINE SIMILARITY PADA SISTEM REKOMENDASI JOB SEEKER

LAPORAN PROGRAM TUGAS AKHIR

Laporan ini dibuat untuk memenuhi persyaratan kelulusan program tugas akhir



Dibuat Oleh:

1.17.4.095

Muhammad Dzihan Albanna

PROGRAM DIPLOMA IV TEKNIK INFORMATIKA

POLITEKNIK POS INDONESIA

BANDUNG

2022

IMPLEMENTATION OF RECOMMENDATION SYSTEM ON JOB RECOMMENDATION BASED ON THE LINKEDIN PROFILE USING COSINE SIMILARITY

INTERNSHIP II PROGRAM REPORT

This report was made to fulfill the requirements for passing the internship II program



Created By:

1.17.4.095

Muhammad Dzihan Albanna

**DIPLOMA IV PROGRAM IN INFORMATION ENGINEERING
POLITEKNIK POS INDONESIA
BANDUNG
2022**

LEMBAR PENGESAHAN

LEMBAR PENGESAHAN

ITEUNG PENERAPAN SISTEM REKOMENDASI PADA JOB RECOMMENDATION BERDASARKAN PROFIL LINKEDIN MENGGUNAKAN COSINE SIMILARITY

Muhammad Dzihan Al-Banna

1.17.4.095

Laporan Program Internship II ini telah diperiksa, disetujui dan disidangkan
Di Bandung, 14 Juli 2022

Oleh :

Ketua Penguji



Roni Habibi, S.Kom., M.T.,SFPC

NIK. 103.78.069

Anggota Penguji



Nisa Hanum Harani, S.Kom.,

M.T.,CDSP.,SFPC

NIK. 117.89.223

Menyetujui,
Koordinator Internship II



Syafrial Fachri Pane, S.T.,M.T.I.,EBDP.,CDSP.,SFPC

NIK. 117.88.233

ITEUNG PENERAPAN SISTEM REKOMENDASI PADA JOB RECOMMENDATION BERDASARKAN PROFIL LINKEDIN MENGGUNAKAN COSINE SIMILARITY

Muhammad Dzihan Al-Banna

1.17.4.095

Laporan Program Internship II ini telah diperiksa, disetujui dan disidangkan

Di Bandung, 14 Juli 2022

Oleh :

Pembimbing Utama,



Roni Habibi, S.Kom., M.T., SFPC

NIK. 103.78.069

Menyetujui,

Ketua Program Studi D4 Teknik Informatika,



M. Yusril Helmi S, S.Kom., M.Kom.

NIK. 113.74.163

LEMBAR BERITA ACARA SIDANG



BERITA ACARA REVISI INTERNSHIP II POLITEKNIK POS INDONESIA TAHUN AJARAN 2021/2022

Saya mahasiswa yang bertanda tangan sebagai berikut,

Nama Mahasiswa : Muhammad Dzihan Al-Banna
NPM : 1174095
Judul : IITEUNG PENERAPAN SISTEM REKOMENDASI PADA JOB RECOMMENDATION BERDASARKAN PROFIL LINKEDIN MENGGUNAKAN COSINE SIMILARITY
Tanggal : 14 Juli 2022
Waktu : 08.00 - 09.00

telah melakukan revisi Internship II pada Prodi D4 Teknik Informatika Politeknik Pos Indonesia Tahun Ajaran 2021/2022 sesuai dengan peraturan yang berlaku, berdasarkan dengan perubahan aturan yang ditetapkan selama masa pandemi Covid 19 dengan kegiatan yang dilaksanakan secara daring dan dokumen elektronik.

Berikut hasil revisi Internship II:

Penguji Utama

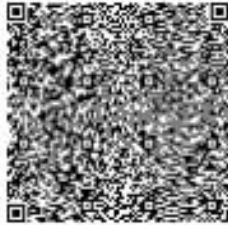
| No | Uraian | Tanda Tangan Mhs |
|----|--|------------------|
| 1 | Semua gambar diperjelas dan berikan penjelasan | |
| 2 | Tulisan di cek lagi spasinya dari awal sampe akhir laporan, bahasa inggris atau bahasa asing dimiringkan | |
| 3 | Kesimpulan menjawab dari tujuan di bab 1 dan jelaskan kesimpulan nya seperti apa | |
| 4 | Saran. | |

Penguji Pendamping

| No | Uraian | Tanda Tangan Mhs |
|----|--|------------------|
| 1 | Buat penjelasan cosine similarity | |
| 2 | Tambahkan rumus cosine similarity di bab teori | |
| 3 | Perbaiki hasil akhir cosine similarity | |
| 4 | Pertimbangkan metode supervised dan unsupervised | |
| 5 | Perbaiki page break | |

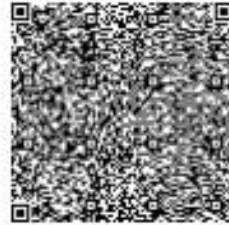
Bandung, 14 Juli 2022Mengetahui:

Penguji Utama



(Roni Habibi, S.Kom., M.T.,SFPC)

Penguji Pendamping

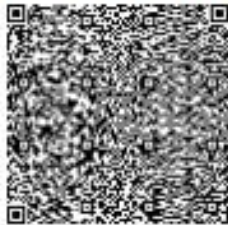


(Nisa Hanum Harani, S.Kom., M.T.,CDSP,SFPC)

Diperiksa oleh,

Ketua

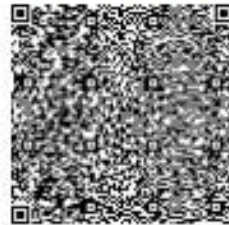
Program Studi D4 Teknik Informatika



(M. Yusril Helmi S, S.Kom., M.Kom.)

Koordinator Internship II

Program Studi D4 Teknik Informatika



(Syafrial Fachri Pane, S.T.,M.T.I.,EBDP,CDSP,SFPC)

ABSTRAK

Perkembangan di bidang teknologi membuat informasi dapat dengan mudah tersebar di berbagai macam platform, informasi yang disajikan tentu bermacam-macam dan disesuaikan dengan kebutuhan. Informasi mengenai pekerjaan selalu menjadi informasi penting yang dicari oleh berbagai kalangan. Sekarang informasi mengenai pekerjaan sudah bertransformasi sehingga lebih mudah dicari hanya dengan mengandalkan browsing saja. Informasi tersebut kini semakin banyak dan tentu sudah banyak juga sistem pemfilteran yang digunakan untuk menyaring informasi salah satunya adalah sistem rekomendasi. Sistem rekomendasi dapat memberikan rekomendasi pekerjaan yang sesuai dengan profil yang dimiliki oleh karyawan pada platform pencarian kerja. Salah satu metode yang digunakan untuk recommendation system adalah content-based filtering yang membandingkan kesamaan fitur. Pada penelitian ini penulis memilih metode *naïve bayes* dan *cosine similarity* sebagai metode yang digunakan untuk membandingkan profil dengan deskripsi pekerjaan sehingga menghasilkan sebuah rekomendasi pekerjaan yang dinilai cocok. Kedua metode tersebut dapat memberikan keluaran yang bisa diterapkan pada *recommendation system* berupa nilai probabilitas dan similarity yang kemudian dilakukan penggabungan atau kombinasi antar kedua metode tersebut untuk mendapatkan akurasi yang lebih maksimal. Hasilnya dari kalkulasi kedua metode tersebut dapat membuat nilai kemiripan antara profil dengan deskripsi pekerjaan menjadi lebih meningkat sehingga dapat dipastikan jarak perbedaannya dengan hasil pencocokan yang lain.

Kata kunci: Sistem Rekomendasi, Content-Based Filtering, Naïve bayes, Cosine Similarity

ABSTRACT

Advances in technology make information can be easily spread across various platforms, the information presented is of course various and tailored to the needs. Information about work has always been important information sought by various groups. Now information about work has been transformed so that it is easier to find just by relying on search alone. The information is now more and more and of course there are also many filtering systems used to filter information, one of which is the recommendation system. The recommendation system can provide job recommendations that match the profiles owned by employees on the job search platform. One of the methods used for the recommendation system is content-based filtering that compares the same features. In this study the authors chose the nave Bayes method and cosine similarity as the method used to compare profiles with job descriptions so as to produce a job recommendation that was considered suitable. Both methods can provide output that can be applied to the recommendation system in the form of probability and similarity values which are then combined or combined between the two methods to obtain maximum accuracy. The results from the calculation of the two methods can increase the value between the profile and the job description so that it can be ascertained the difference between the results of other things.

Keyword : Recommendation System, Content-Based Filtering, Naïve Bayes, Cosine Similarity

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh. Puji syukur penulis panjatkan ke hadirat Allah SWT karena hanya dengan rahmat dan hidayahnya, laporan internship II ini dapat terselesaikan tanpa halangan berarti. Keberhasilan dalam menyusun laporan internship II ini tidak lepas dari bantuan berbagai pihak yang mana dengan tulus dan ikhlas memberikan masukan guna sempurnanya laporan internship II ini. Oleh karena itu, dalam kesempatan ini, dengan kerendahan hati penulis mengucapkan terima kasih kepada

1. M. Yusril Helmi Setyawan, S.Kom., M.Kom. Selaku Ketua Program Studi DIV Teknik Informatika Politeknik Pos Indonesia,
2. Syafrial Fachri Pane, ST., MTI., EBDP., CDSP., SFPC Selaku Koordinator internship II
3. Orang Tua, yang selalu memberi dukungan dan doa,
4. Teman - teman yang juga memberi dukungan dan doa.

Penulis menyadari bahwa penyusunan laporan internship II ini jauh dari sempurna. Kritik dan saran dapat ditujukan kontak penulis. Dengan terselesaikannya laporan internship II, penulis berharap semoga laporan ini dapat menimbulkan manfaat bagi para pembaca.

Wassalamualaikum Wr. Wb.

Bandung, 06 Mei 2022

Penulis

DAFTAR ISI

| | |
|---|---|
| LEMBAR PENGESAHAN | i |
| LEMBAR BERITA ACARA SIDANG | i |
| ABSTRAK..... | iii |
| <i>ABSTRACT</i> | iv |
| KATA PENGANTAR | v |
| DAFTAR ISI..... | vi |
| DAFTAR GAMBAR..... | viii |
| DAFTAR TABEL..... | ix |
| DAFTAR SIMBOL | x |
| DAFTAR SINGKATAN | xi |
| DAFTAR LAMPIRAN..... | xii |
| RINGKASAN | xiii |
| BAB I PENDAHULUAN..... | I-1 |
| 1.1 Latar Belakang | I-1 |
| 1.2 Identifikasi Masalah | I-3 |
| 1.3 Tujuan dan Manfaat | I-3 |
| 1.4 Ruang Lingkup..... | I-3 |
| 1.5 Sistematika penulisan..... | I-4 |
| BAB II TINJAUAN PUSTAKA | II-5 |
| 2.1 State of The Art (SOTA)..... | II-5 |
| 2.2 Landasan Teori..... | II-6 |
| 2.3 Tinjauan Pustaka | II-11 |
| BAB III GAMBARAN UMUM PERUSAHAAN | III-Error! Bookmark not defined. |
| 3.1 Struktur Organisasi..... | III-Error! Bookmark not defined. |
| 3.2 Business Plan..... | III-Error! Bookmark not defined. |
| 3.3 Cash Flow..... | III-Error! Bookmark not defined. |
| 3.4 <i>Job Description</i> Peserta Internship | III-Error! Bookmark not defined. |
| 3.5 Ruang Lingkup Peserta Internship | III-Error! Bookmark not defined. |
| BAB IV METODE PENELITIAN | IV-14 |
| 4.1 Diagram Alur Metodologi Penelitian..... | IV-14 |
| 4.2 Indikator Capaian Penelitian | IV-20 |

| | |
|---|--------------------------------|
| BAB V SIMULASI DATA | V-21 |
| 5.1 Data yang Digunakan | V-Error! Bookmark not defined. |
| 5.2 Data User..... | V-Error! Bookmark not defined. |
| 5.3 Data Jobs | V-Error! Bookmark not defined. |
| 5.4 Simulasi Data | V-Error! Bookmark not defined. |
| 5.5 <i>Prototype</i> | V-Error! Bookmark not defined. |
| BAB VI PENUTUP | VI-28 |
| 6.1 Kesimpulan dan Saran..... | VI-28 |
| DAFTAR PUSTAKA | 29 |
| LAMPIRAN-LAMPIRAN | 33 |
| Lampiran 1. Bukti Kartu Bimbingan (KAMBING) | 33 |
| Lampiran 2. Bukti Pengecekan Plagiarisme Online | 35 |
| Lampiran 3. Surat Pernyataan..... | 36 |
| Lampiran 4. Draft Jurnal..... | 38 |



DAFTAR GAMBAR

| | |
|--|----------------------------------|
| Gambar 3. 1 Struktur Organisasi | III-Error! Bookmark not defined. |
| Gambar 3. 2Bisnis Model Canvas | III-Error! Bookmark not defined. |
| Gambar 4. 1 Gambar Diagram Alur Metodologi Penelitian..... | IV-14 |
| Gambar 5. 1 Simulasi Proses Scraping Data Jobs | V-Error! Bookmark not defined. |
| Gambar 5. 2 Simulasi Proses Scraping Data User | V-Error! Bookmark not defined. |
| Gambar 5. 3 Simulasi Proses Pembersihan Data | V-Error! Bookmark not defined. |
| Gambar 5. 4 Simulasi Proses Word Embedding..... | V-Error! Bookmark not defined. |
| Gambar 5. 5 Simulasi Proses Feature Extraction..... | V-Error! Bookmark not defined. |
| Gambar 5. 6 Simulasi Proses Sistem Rekomendasi..... | V-Error! Bookmark not defined. |
| Gambar 5. 7 Prototype Sistem Rekomedasi | V-Error! Bookmark not defined. |
| Gambar 5. 8 Prototype Hasil Rekomendasi | V-Error! Bookmark not defined. |

DAFTAR TABEL

| | |
|--|--|
| Tabel 2. 1 Penelitian Terkait..... | II-11 |
| Tabel 3. 1 Cash Flow | III- Error! Bookmark not defined. |
| Tabel 4. 1 Indikator Capaian Penelitian..... | IV-20 |

DAFTAR SIMBOL

| | | |
|----|---|--|
| 1. |  | Arrows, menunjukkan arus data antar simbol atau proses. |
| 2. |  | Menunjukkan kegiatan atau proses dari operasi program komputer. Menggambarkan setiap proses dalam <i>flowmap</i> dari awal hingga akhir. |

DAFTAR SINGKATAN

| No. | Singkatan | Arti |
|-----|-----------------|--|
| 1. | <i>CRISP-DM</i> | <i>CRoss-Industry Standard Process for Data Mining</i> |
| 2. | <i>HR</i> | <i>Human Resource</i> |
| 3. | <i>CF</i> | <i>Collaborative filtering</i> |

DAFTAR LAMPIRAN

| | |
|--|----|
| Lampiran 1 Lembar Kartu Bimbingan (KAMBING)..... | 33 |
| Lampiran 2 Lembar Bukti Pengecekan Plagiarisme Online..... | 35 |
| Lampiran 3 Surat Pernyataan..... | 36 |
| Lampiran 4 Draft Jurnal..... | 38 |

RINGKASAN

BAB I

PENDAHULUAN

1.1 Latar Belakang

Perkembangan di bidang teknologi membuat informasi dapat dengan mudah tersebar di internet, begitu pula dengan lowongan pekerjaan yang sudah bertransformasi menjadi lebih mudah dicari dengan mengandalkan browsing saja. Saat mencari pekerjaan tentu akan tersedia banyak sekali informasi yang tersedia di internet, untuk memudahkan saat mencari pekerjaan dibutuhkan sebuah sistem yang bisa merekomendasikan item tertentu yang sesuai dengan pengguna, sistem tersebut dinamakan *recommendation system*(Rahmawati et al., 2018)

Untuk membuat sistem yang dapat merekomendasikan item yang sesuai dengan pengguna maka dibutuhkan bidang yang khusus membahas pengetahuan mengenai data dan membuat sebuah informasi dari data yang kompleks, bidang ilmu tersebut adalah data science (Sanchez-Pinto et al., 2018). Tujuan dari ilmu data itu sendiri adalah mengembangkan sebuah metode untuk melakukan ekstraksi ilmu dan pengetahuan yang berasal dari kumpulan data. Kemampuan dalam mengolah data belakangan sudah meningkat dengan signifikan sehingga untuk memahami data yang banyak diperlukan ketelitian dan membutuhkan cara yang lebih efektif dalam mengekstrak informasi. Untuk mengatasi hal tersebut maka adanya *machine learning* akan membantu dalam melakukan tugas yang eksekusinya akan memakan waktu yang lama(Ruehle, 2020). Ilmu data lebih dari sekedar menganalisis data. Banyak perangkat lunak yang berbeda digunakan untuk mengatur, mengumpulkan, memvisualisasikan, dan menyajikan data. Akhirnya, banyak sistem yang berbeda terlibat dalam pengerjaan data(Govindarajan, 2020). Dalam masalah sehari-hari data science dapat memberikan sebuah rekomendasi seperti rekomendasi pekerjaan ataupun memberikan rekomendasi film, jurusan, makanan dan lain-lain(Kenthapadi et al., 2017)(D. C. G. Putri et al., 2020). Data science telah berkembang menjadi ilmu yang fleksibel sehingga sering digunakan untuk membantu kebutuhan bisnis seperti menyiapkan strategi perusahaan(Husein & Harahap, 2021). Dalam data science terdapat pendekatan *machine learning* yang membuat sistem belajar sendiri sehingga dapat membuat keputusan sendiri(Retnoningsih & Pramudita, 2020) terdapat beberapa pendekatan yang ada dalam *machine learning* diantaranya *supervised learning*, *unsupervised*, *semi-supervised* dan *reinforcement learning*.(Portugal et al., 2018) Dalam algoritma *unsupervised learning* disajikan data mentah tanpa adanya set training sehingga diharuskan untuk bisa belajar dari data itu sendiri. Sedangkan pendekatan *semi-supervised learning* digunakan ketika ada

informasi yang hilang pada algoritma yang menggunakan training set dan masih membutuhkan bantuan dan gambaran mengenai data yang belum lengkap. Misalnya, untuk rating film, tidak setiap pengguna menilai setiap film dan ada beberapa informasi yang hilang. Algoritma *semi-supervised learning* mampu mempelajari dan menarik kesimpulan bahkan dengan data yang tidak lengkap (Portugal et al., 2018).

Dengan munculnya beragam opsi dalam pengolahan data membuat kinerja manusia semakin berkurang (Nasution, 2021). Sama halnya dengan kemunculan sistem rekomendasi dalam sebuah sistem yang membuat penggunaannya tidak perlu bertanya kepada manusia tentang rekomendasi item yang cocok untuknya (Badriyah et al., 2018). Sistem rekomendasi merupakan program yang dapat bertindak sebagai perantara membuat daftar informasi yang dibutuhkan dan membandingkan kecocokan antara item dengan pengguna (Prasetya, 2017). Skema rekomendasi kini menjadi sangat penting di berbagai bidang, termasuk *e-commerce*, jaringan sosial, dan sejumlah layanan berbasis web (J. Zhang et al., 2020). Tidak hanya memberikan rekomendasi item dalam *e-commerce* sistem rekomendasi juga dapat memberikan rekomendasi pekerjaan bagi *jobseeker* yang bertujuan untuk memberikan rekomendasi pekerjaan yang sesuai dengan keahlian (Rahmawati et al., 2018).

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 (Grashinta et al., 2018). Metode yang akan digunakan adalah *content-based* yang dilakukan dengan menghitung kemiripan antara konten (Mustika, 2017). Disini *user* sudah memiliki profil tetap untuk mendeskripsikan dirinya, kemudian sistem memberikan item yang relevan dengan profil masing-masing *user*, sistem akan memilih dan membuat *ranking* pada item yang sesuai setelah itu akan dikirimkan ke pengguna (Nastiti, 2019). Selain itu dalam penelitian ini akan menggunakan pendekatan *Cosine Similarity* untuk membandingkan kedekatan antar teks atau *text matching* (Suka Parwita et al., 2018). *Text matching* atau Pencocokan teks adalah salah satu masalah penelitian terpenting di banyak kasus seperti informasi, jawaban pertanyaan, dan rekomendasi (J.-Y. Jiang et al., 2019). *Cosine Similarity* adalah *similarity* paling populer yang digunakan di banyak website *e-commerce* termasuk Amazon dan YouTube (Bag et al., 2019). Selain itu untuk memaksimalkan hasil

rekomendasi pada penelitian ini akan digunakan juga metode Naïve Bayes. Naive Bayes adalah metode probabilistik untuk klasifikasi. Metode ini digunakan untuk menentukan probabilitas dari contoh kelas yang diberikan nilai dari variable(Intan Aprilia & Fachrurrozi, 2016). Pada penelitian sebelumnya menggunakan *cosine similarity* menghasilkan rekomendasi pekerjaan dengan nilai similarity tertinggi yang dijadikan acuan sebagai pekerjaan yang dianggap sesuai dengan *user*. Adanya penelitian lanjutan dari penelitian sebelumnya diharapkan dapat memberikan rekomendasi yang dapat dilihat akurasi dan juga melengkapi hasil rekomendasi yang masih kurang sesuai. Penelitian ini Hasil dari sistem rekomendasi pekerjaan akan ditampilkan dalam bentuk web menggunakan *Flask*.

1.2 Identifikasi Masalah

Adapun identifikasi masalah pada penelitian ini sebagai berikut,

- a) Bagaimana memberikan rekomendasi pekerjaan yang sesuai bagi *jobseeker*?
- b) Bagaimana menampilkan hasil dari rekomendasi pekerjaan yang mudah dipahami?

1.3 Tujuan dan Manfaat

Adapun tujuan pada penelitian ini sebagai berikut,

- a) Memberikan rekomendasi pekerjaan yang sesuai bagi *jobseeker* berdasarkan profil.
- b) Memberikan hasil rekomendasi pekerjaan menggunakan flask.

Adapun manfaat pada penelitian ini sebagai berikut,

- a) Dapat memberikan rekomendasi pekerjaan yang relevan berdasarkan profil *linkedin*.
- b) Dapat memberikan peringkat kecocokan antara *skill* dengan *job desc*.

1.4 Ruang Lingkup

Adapun ruang lingkup pada penelitian ini sebagai berikut,

- a) Data yang digunakan untuk pengujian adalah data dari user *linkedin* yang sudah dikumpulkan melalui *google form*.
- b) Pengujian hanya dilakukan pada data yang didapatkan saat penyebaran *google form*.
- c) Penelitian hanya dilakukan pada lingkungan tempat peneliti melakukan Internship II.

1.5 Sistematika penulisan

Berdasarkan latar belakang dan perumusan masalah diatas, maka penyusunan laporan ini dibuat dalam suatu sistematika yang terdiri dalam empat BAB, yaitu:

BAB I PENDAHULUAN

Pada bab ini dijelaskan latar belakang, identifikasi masalah, tujuan, ruang lingkup dan sistematika penulisan.

BAB II LANDASAN TEORI

Pada bab ini dijelaskan teori-teori dan penelitian terdahulu yang digunakan sebagai acuan dan dasar dalam penelitian.

BAB III METODE PENELITIAN

Pada bab ini berisi tentang jadwal penelitian.

BAB IV HASIL IMPLEMENTASI

Pada bab ini menjelaskan dari hasil pengkajian sementara dan simulasi beserta prototype

BAB V PENUTUP

Pada bab terakhir memaparkan mengenai kesimpulan dari penelitian

BAB II

LANDASAN TEORI

2.1 State of The Art (SOTA)

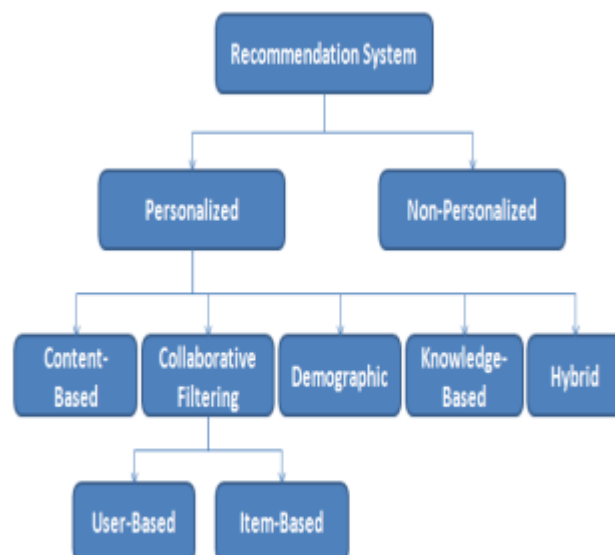
Kemajuan teknologi pada revolusi industri 4.0 membuat perubahan di berbagai macam sektor yang diubah menjadi serba otomatis(Nasution, 2021). Namun dalam sebuah teknologi akan terus berkembang seiring berjalannya waktu dikarenakan muncul kembali beberapa masalah yang dihadapi oleh manusia salah satunya adalah kesulitan disaat mencari pekerjaan. Banyak pelamar yang mencari pekerjaan menggunakan situs lowongan kerja secara online (Rahmawati et al., 2018). Namun dikarenakan sangat banyak pelamar dan pekerjaan yang tersedia maka terdapat kesulitan disaat mencari pekerjaan yang sesuai dengan keahlian yang dimiliki pelamar(Rahmawati et al., 2018). Maka dibutuhkan sebuah sistem rekomendasi yang dapat membuat beberapa kombinasi item yang akan memberikan hal-hal yang dianggap sesuai dengan *user*(Badriyah et al., 2018). Untuk mencapai tujuan tersebut diperlukan salah satu bidang pekerjaan yang khusus membahas ekstraksi data yang menjadikan sebuah data menjadi informasi yang bisa dipahami(Ruehle, 2020). Dengan bantuan sistem rekomendasi para pelamar kerja akan diberikan rekomendasi pekerjaan yang sesuai dengan keahliannya(Rahmawati et al., 2018). Untuk mendukung tujuan tersebut peneliti memilih metode *content-based filtering* karena user memiliki profil konsisten yang diambil dari LinkedIn untuk kemudian diberikan rekomendasi berupa job yang relevan dengan profilnya, kemudian akan dilakukan urutan peringkat dari *similarity* terbesar hingga terkecil(Nastiti, 2019). Untuk membandingkan kemiripan antar teks digunakan pendekatan *Cosine Similarity* dan *flask* sebagai framework untuk visualisasi dalam bentuk web(Suka Parwita et al., 2018). Selain itu, tambahan metode Naïve Bayes bertujuan untuk membuat hasil rekomendasi yang lebih spesifik, pemilihan metode Naïve Bayes digunakan karena salah satu keunggulan penggunaan algoritma Naive Bayes Classifier adalah nilai keputusan yang dihasilkan pada nilai probabilistik dengan independensi (ketidak bergantungan) yang kuat(Asfi & Fitrianiingsih, 2020). Metode ini hanya membutuhkan jumlah data pelatihan (training data) yang sedikit untuk menentukan estimasi parameter yang diperlukan dalam proses pengklarifikasian dan waktu training yang lebih cepat dibandingkan metode lainnya(Intan Aprilia & Fachrurrozi, 2016).

Peneliti telah mengkaji beberapa referensi yang berkaitan dengan sistem rekomendasi dan beberapa hal yang merupakan manfaat dan kekurangan dari sistem rekomendasi yang telah ada sebelumnya. Referensi yang telah dikaji ditunjukkan pada tabel 2.1 :

2.2 Landasan Teori

2.2.1 Sistem Rekomendasi

Sistem rekomendasi merupakan program yang dapat bertindak sebagai perantara membuat daftar informasi yang dibutuhkan dan membandingkan kecocokan antara item dengan pengguna (Prasetya, 2017). Skema rekomendasi kini menjadi sangat penting di berbagai bidang, termasuk *e-commerce*, jaringan sosial, dan sejumlah layanan berbasis web (J. Zhang et al., 2020). Contohnya sistem rekomendasi film yang bisa memprediksi film apa yang akan disukai pengguna berdasarkan atribut yang ada di film yang sebelumnya disukai (Reddy et al., 2019). Sebagian *Movie recommendation* menggunakan model *Collaborative filtering* (CF) dan *Content-based Filtering* seperti genre atau pemeran (Deldjoo et al., 2019). Sistem rekomendasi biasanya melibatkan pemanfaatan hubungan antara fitur dan konten yang diketahui yang menggambarkan item pemfilteran berbasis konten (*Content-based Filtering*) atau tumpang tindih pengguna serupa yang berinteraksi atau memberi peringkat (*Collaborative filtering*) (Yang et al., 2017).



Dalam *e-commerce* retail sistem rekomendasi juga digunakan untuk merekomendasikan sesuatu yang mungkin dia sukai karena pernah membeli atau ‘menyukai’ item tersebut (Hwangbo et al., 2018). Karena konsumen ingin menemukan item favorit dengan cepat

dari sekian banyaknya barang yang terdapat di *e-commerce* (L. Jiang et al., 2019). Selain bidang-bidang tersebut pada sistem pencarian kerja berbasis web yang semakin banyak pelamar yang mencari lowongan pekerjaan setiap hari dan berurusan dengan perekrutan yang sangat besar informasi di Internet, pencari kerja selalu menghabiskan waktu berjam-jam untuk menemukan pekerjaan yang dicari membuat munculnya ide untuk mengadakan rekomendasi pekerjaan pada online *job boards* membuat rekomendasi yang akurat dibutuhkan (Desai et al., 2017) (Shalaby et al., 2017).

2.2.2 Word Embedding

Word Embedding adalah representasi vektor bernilai dengan menyematkan makna semantik dan sintaksis diperoleh dari *corpus* besar yang tidak berlabel (Wang et al., 2019). *Word embedding* terbagi lagi menjadi beberapa bagian yaitu *word2vec*, *glove*, *fasttext* (Nurdin et al., 2020). Model *Word2vec* melakukan proses training dengan cara memprediksi kata sebelumnya dan juga kata sesudahnya berdasarkan kata saat ini (Rahman et al., 2021). Model *GloVe* telah menunjukkan hasil yang sukses dalam mengubah data tekstual menjadi vektor fitur (Gautam & Jerripathula, 2020). *GloVe* dinilai lebih efisien daripada *Word2Vec*, karena vektor dimensi *word2vec* rendah, dan tidak dapat memasukkan semua informasi dari *corpus* (Mohammed et al., 2021). Namun dalam penelitian lain secara umum, *word embedding FastText* dengan 300 dimensi melaporkan akurasi terbaik di semua tugas evaluasi, sementara *Glove* melaporkan hasil terendah (Lakmal et al., 2020).

2.2.3 Glove

Algoritma *GloVe* merupakan salah satu metode representasi vektor kata. Algoritma ini menggabungkan keunggulan suatu text dan faktorisasi matriks global. *GloVe* memiliki banyak *hyperparameter* itu dapat mempengaruhi hasilnya, tetapi ukuran dan dimensi vektor adalah yang paling efektif (Ibrahim et al., 2020). Mendefinisikan kata sebagai vektor memudahkan algoritme *machine learning* untuk memahami teks dan mengekstraknya (Ibrahim et al., 2021).

2.2.4 Content Based-filtering

Sistem Rekomendasi dibagi kepada beberapa bagian yaitu *Personalized* dan *NonPersonalized*. Untuk *personalized* itu sendiri dibagi menjadi *Content-based Filtering*, *Collaborative filtering*, *Hybrid*, *Knowledge-based*, dan *Demographic* (Das et al., 2017). Masing-masing memiliki kegunaan tersendiri seperti *collaboratif filtering* yang mampu menangkap interaksi atau korelasi pengguna dan item yang sedang dipertimbangkan (Fu et al., 2019) seperti penilaian

kepada beberapa item pilihan user dan juga yang dilakukan oleh user lain yang selanjutnya akan dijadikan sebagai acuan untuk memberikan prediksi terhadap pengguna lain atau pengguna yang berkaitan (Simbolon & Kasih, 2019), *Collaborative Filtering* mengolah data dengan menyaring berdasarkan similarity karakteristik dari setiap pengguna (Nugroho, 2020). *Collaborative Filtering* dibagi menjadi dua, yaitu *Item-based Collaborative Filtering*, dan *User-based Collaborative Filtering* (Jepriana & Wardoyo, 2018). Perbedaan dari kedua metode tersebut adalah *Item-based Collaborative Filtering* beranggapan jika semakin banyak user yang melakukan pemilihan dan memberi ulasan, maka user selanjutnya akan memberikan ulasan juga (Ma et al., 2017) sedangkan *User-based Collaborative Filtering* menilai bahwa jika user memiliki kecenderungan atau kemiripan yang sama maka akurasi akan semakin tinggi (Z. Zhang et al., 2017) yang serupa terhadap item tersebut. *Hybrid* adalah gabungan dari beberapa metode pada sistem rekomendasi (Wijaya & Alfian, 2018) sedangkan *content-based filtering* dilakukan dengan menghitung kemiripan antara konten (Mustika, 2017). Metode *Content-Based Filtering* bekerja dengan melihat kemiripan item baru dengan item yang sebelumnya. *Content-Based Filtering* memberikan rekomendasi berdasarkan kemiripan item yang dianalisis dari fitur yang dikandung oleh item sebelumnya seperti deskripsi, sinopsis atau genre item yang mana buku pada perpustakaan juga terdiri dari genre atau kategori tertentu (Alkaff et al., 2020). Sistem rekomendasi menampilkan hasil rekomendasi produk untuk pengguna yang sesuai atau memiliki kesamaan antara atribut konten dan atribut yang dimiliki oleh pengguna (M. W. Putri et al., 2020).



2.2.5 Cosine Similarity

Cosine Similarity adalah *similarity* paling populer yang digunakan di banyak website *e-commerce* termasuk Amazon dan YouTube (Bag et al., 2019). *Cosine Similarity* merupakan algoritma dimana nilai *similarity* antara i dan j digambarkan sebagai suatu sudut yang terbentuk diantara 2 buah (Ferio et al., 2019). Secara umum sistem yang dibangun menggunakan metode *Cosine Similarity* yaitu untuk menghitung *similarity* (tingkat kesamaan) antar dua buah objek misalnya $D1$ dan $D2$ yang dinyatakan dalam dua buah vector (Yasni et al., 2018). *cosine similarity* berfungsi untuk menguji ukuran yang dapat digunakan sebagai interpretasi dari

kedekatan jarak berdasarkan kesamaan menggunakan dataset yang ada. Kemiripan antara dua item, kebiasaan, dan ditentukan oleh jumlah kumpulan kejadian dari dua item di sesi lain. Dengan demikian, kesamaan antar item kemudian dapat dihitung (Budiman & Mastrika Giri, 2020).

$$\cos(x, y) = \frac{x \cdot y}{||x|| * ||y||}$$

Contoh :

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

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

Rumus cosin similarity adalah :

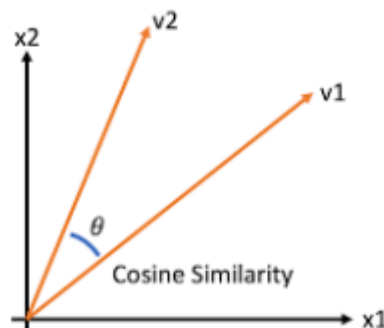
$$\cos(x, y) = \frac{x \cdot y}{||x|| * ||y||}$$

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

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

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

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



Kemiripan yang digunakan sebagai parameter oleh *Cosine Similarity* diambil dari sudut cosinus dari dua objek yang berbeda yang kemudian dibandingkan antara kedua objek tersebut. Sudut theta antara dua film akan menentukan kesamaan antara kedua film tersebut. Theta berkisar dari 0- 1. Jika nilai theta mendekati 1 maka paling mirip dan jika mendekati 0 maka paling tidak mirip (Singh et al., 2020)

2.2.6 Naïve Bayes

Naive Bayes adalah metode probabilistik untuk klasifikasi. Metode ini digunakan untuk menentukan probabilitas dari contoh kelas yang diberikan nilai dari variabel (Intan Aprilia &

Fachrurrozi, 2016). *Naïve bayes* dapat memberikan akurasi klasifikasi yang kompetitif. Ditambah dengan efisiensi komputasinya dan banyak fitur lain yang diinginkan, ini menyebabkan *Naïve Bayes* diterapkan secara luas di praktek (Webb, 2020). Algoritma *naïve bayes* dapat digunakan untuk klasifikasi dengan data bertipe nominal dan juga bertipe numerik. Data bertipe nominal yaitu tipe data yang didapat dari proses klasifikasi atau kategorisasi. Sedangkan tipe data bertipe numerik adalah data yang didapat dari proses pengukuran seperti umur, berat badan dan tinggi badan. Selain daripada itu dalam penerapannya *naïve bayes* memiliki beberapa pilihan seperti metode multinomial yang biasa digunakan untuk klasifikasi dokumen, dimana fitur atau prediktor yang digunakan oleh pengklasifikasi adalah frekuensi dari kata-kata yang ada dalam dokumen. *Naïve Bayes* menjadi salah satu algoritma yang biasa digunakan untuk membuat sistem rekomendasi dengan mengandalkan perhitungan probabilitas antar kriteria yang sebelumnya sudah ditetapkan. Dalam beberapa penelitian yang telah dilakukan *Naïve Bayes* dapat memberikan rekomendasi yang digunakan untuk system rekomendasi film, jurusan perkuliahan, dan juga pembimbing skripsi (Asfi & Fitrianiingsih, 2020). Rumus umum pada *naïve bayes* adalah sebagai berikut.

$$P(H|X) = \frac{p(X|H) \cdot p(H)}{p(X)}$$

X : Data dengan class yang belum diketahui

H : Hipotesis data merupakan suatu class spesifik

$P(H|X)$: Probabilitas hipotesis H berdasar kondisi X (posteriori probabilitas)

$P(H)$: Probabilitas hipotesis H (prior probabilitas)

$P(X|H)$: Probabilitas X berdasarkan kondisi pada hipotesis H

$P(X)$: Probabilitas X

2.3 Tinjauan Pustaka

Tabel 2. 1 Penelitian Terkait

| No. | Area Penelitian | Tahun | Karakteristik Data | Metode | Hasil Penelitian |
|-----|--|-------|--|--------------------------------|---|
| 1. | Sistem Rekomendasi <i>Content Based Filtering</i> Menggunakan Algoritma Apriori(Badriyah et al., 2018) | 2018 | Produk <i>e-commerce</i> . | <i>Content-based filtering</i> | Menghasilkan sistem rekomendasi pada <i>e-commerce</i> , Nilai <i>support</i> dan <i>confidence</i> mempengaruhi jumlah rekomendasi |
| 2 | <i>Design of an Unsupervised Machine Learning-Based Movie Recommender System</i> (D. C. G. Putri et al., 2020) | 2020 | <i>MovieLans dataset, genre, tag, rating</i> | <i>Unsupervised Learning</i> | Pengelompokan dioptimalkan yang dikembangkan dari beberapa algoritma adalah kemudian digunakan untuk membandingkan algoritme terbaik sehubungan dengan pengelompokan kesamaan pengguna di film genre, tag, dan peringkat menggunakan dataset <i>MovieLens</i> |
| 3 | Penerapan Metode <i>Content Based Filtering</i> Dalam Implementasi | 2019 | Data Tanaman Pangan, petani, pedagang | <i>Content-based filtering</i> | Petani dapat menemukan pedagang yang menjual/membeli tanaman pangan yang sama dengan yang |

| | | | | | |
|---|---|------|---|---|---|
| | Sistem Rekomendasi Tanaman Pangan(Nastiti, 2019) | | | | ditanam petani. berdasarkan pengujian terhadap 10 profil pedagang dengan 15 rekomendasi kelompok tani teratas, didapatkan hasil presisi rata-rata sebesar 78.40% |
| 4 | <i>Combining content-based and collaborative filtering for job recommendation system: A cost-sensitive Statistical Relational Learning approach</i> (Yang et al., 2017) | 2017 | <i>Job, User</i> | <i>Content-based filtering, Collaborative Filtering</i> | kemampuan model dapat mengurangi tingkat rekomendasi pekerjaan yang tidak sesuai |
| 5 | Implementasi <i>Cosine Similarity Matching</i> Dalam Penentuan Dosen Pembimbing Tugas Akhir (Yasni et al., 2018) | 2018 | Data Dosen Pembimbing, bidang keahlian, abstrak | <i>Cosine Similarity</i> | <i>Cosine similarity</i> dapat diterapkan pada sistem yang menjadi media rekomendasi, terbukti dengan pengujian sistem yang dilakukan memperoleh performa yang cukup baik yaitu hasil pengujian <i>precision</i> dan <i>recall</i> memiliki |

| | | | | | |
|--|--|--|--|--|----------------------------------|
| | | | | | performa rata-rata 0.74 dan 1 |
|--|--|--|--|--|----------------------------------|

BAB III

METODE PENELITIAN

4.1 Diagram Alur Metodologi 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 (Martinez-Plumed et al., 2021).



Gambar 4. 1 Gambar Diagram Alur Metodologi Penelitian

1. *Bussiness Understanding*

Business understanding adalah pemahaman terhadap tujuan bisnis, penilaian situasi dan menerjemahkan tujuan bisnis ke dalam tujuan *data mining*. Dalam internship ini membutuhkan pengetahuan untuk mendapatkan data dari *course-course* yang diadakan secara online, *keyword* dari paper dan topik dari setiap referensi *course*. Tidak lupa pula melakukan analisa untuk mendukung proses rekomendasi pelatihan menggunakan dataset *PT. XYZ*. 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.

2. *Data Understanding*

Pada tahap ini dilakukan proses pengumpulan data, kemudian melakukan analisa data serta melakukan evaluasi kualitas data yang digunakan dalam pengerjaan internship. Sumber data yang digunakan dalam internship ini merupakan data-data yang terdapat dari berbagai media. Maka sumber literatur banyak didapatkan dari buku, paper atau jurnal, karya ilmiah dan situs-situs penunjang. 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*.

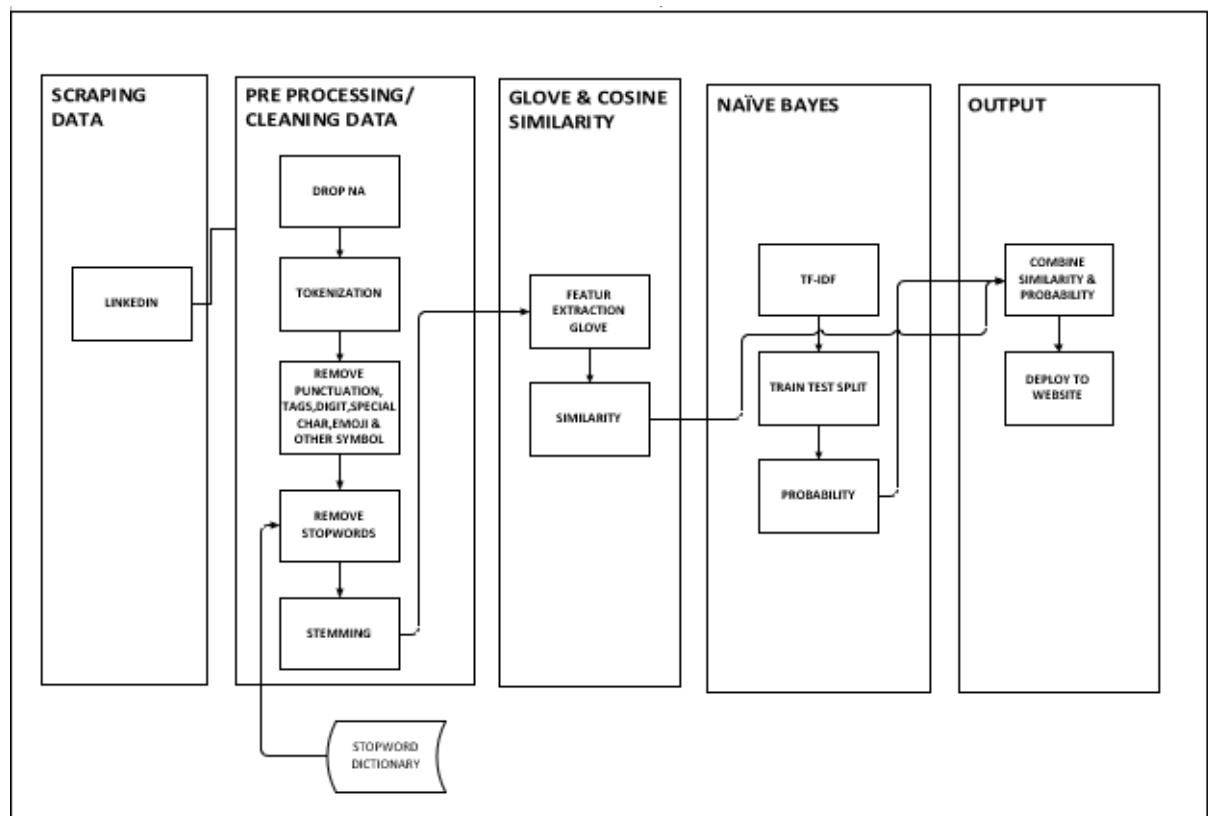
3. *Data Preparation*

Data Preparation atau bisa disebut juga dengan data *preprocessing* adalah suatu proses/langkah yang dilakukan untuk membuat data mentah menjadi data yang berkualitas (input yang baik untuk *data mining tools*). Dengan seperti ini proses pengerjaan internship akan lebih mudah. 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*.

4. *Modelling*

Dalam tahap pemodelan digunakan berbagai macam metode dipilih untuk diterapkan ke dalam dataset yang telah disiapkan. Metode yang digunakan antara lain *naïve bayes* dan *cosine similarity*, teknik yang digunakan adalah *semi-supervised learning*. Pendekatan

semi-supervised learning digunakan ketika ada informasi yang hilang pada algoritma yang menggunakan training set dan masih membutuhkan bantuan dan gambaran mengenai data yang belum lengkap. Misalnya, untuk *rating* film, tidak setiap pengguna menilai setiap film dan ada beberapa informasi yang hilang. Algoritma *semi-supervised learning* mampu mempelajari dan menarik kesimpulan bahkan dengan data yang tidak lengkap. *Semi-supervised learning* berada diantara *supervised learning* dan *unsupervised learning*. Sesuai dengan teknik yang digunakan maka peneliti melakukan tahap *unsupervised* terlebih dahulu dengan metode *similarity*. Setelah itu dilakukan tahap *supervised learning* menggunakan algoritma *naïve bayes*. Untuk mencapai tujuan tersebut terdapat tahapan-tahapan lain yang dilakukan terlebih dahulu seperti *scraping*, *pre processing* dan *cleaning data*. Langkah selanjutnya setelah proses-proses tersebut adalah tahap *deploy* yang akan dilakukan menggunakan *framework flask*. Adapun tahapan-tahapan yang akan dilakukan pada penelitian ini dijelaskan seperti pada gambar berikut.



Pada gambar 3.1 dapat dilihat alur yang akan dilakukan pada penelitian ini.

1. Scraping data

Pada tahap ini dilakukan pengumpulan data dari *platform linkedin*. Dalam melakukan *scrapping* data pada platform linkedin pada prosesnya dibantu menggunakan *library selenium*. Data yang dikumpulkan adalah data *user linkedin* berupa deskripsi *user*.

2. Preprocessing/cleaning data

Tahap ini merupakan pembersihan data, tujuannya adalah agar data yang masih mentah dapat dimengerti dan dapat diimplementasikan menggunakan *data maining tools*. Langkah awal yang dilakukan pada tahap ini sesuai dengan alur diagram diatas yaitu :

a. DROP NA

Proses pemberishan data yang dilakukan di awal adalah *DROP NA*, *NA* merupakan data kosong pada salah satu kolom di dataset yang tidak terisi, data yang tidak terisi tersebut akan dihilangkan.

b. Tokenization

Dalam *tokenization* teks yang awalnya berbentuk kalimat akan diubah menjadi token-token terpisah.

c. Remove Punctuation, Tags, Digits, Special Chars, Emoji & Other Symbol

Pada tahap ini akan dilakukan pembersihan terhadap tanda baca seperti *tags, digits, special chars, emoji* dan *symbol* lainnya.

d. Remove Stopwords

Remove Stopwords berfungsi untuk menghilangkan kata-kata yang sering digunakan namun dianggap tidak penting untuk klasifikasi teks, hal ini bertujuan untuk mengurangi ukuran korpus namun tetap mengutamakan informasi penting.

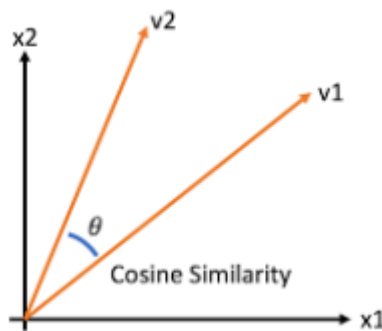
e. Stemming

Stemming digunakan untuk mengambil kata dasar dan menghilangkan kata akhiran.

3. Glove & Cosine Similarity

Tahap ini adalah tahap untuk mencapai sebuah *similarity* yang dimulai dengan tahap *feature extraction* menggunakan metode *glove*, mendefinisikan kata sebagai vektor memudahkan algoritme *machine learning* untuk memahami teks dan mengekstraknya. Dataset yang digunakan diubah menjadi bentuk vector sebagai langkah lanjutan untuk

mempermudah tahap pencocokan menggunakan *cosine similarity*. *Cosine similarity* menggambarkan nilai a dan b seperti sebuah sudut yang terbentuk antara a dan b, metode ini berfungsi untuk menghitung kedekatan jarak berdasarkan kemiripan, kebiasaan dan ditentukan oleh kumpulan kejadian dua item dari dataset yang ada. Maka dari itu kesamaan antara dua buah item dapat dihitung.



Kemiripan yang digunakan sebagai parameter oleh *Cosine Similarity* diambil dari sudut cosinus dari dua objek yang berbeda yang kemudian dibandingkan antara kedua objek tersebut. Sudut theta antara dua film akan menentukan kesamaan antara kedua film tersebut. Theta berkisar dari 0- 1. Jika nilai theta mendekati 1 maka paling mirip dan jika mendekati 0 maka paling tidak mirip. Hasil dari *cosine similarity* dapat menyimpulkan bahwa a dan b memiliki kemiripan dengan acuan jika angka mendekati 1 maka a dan b dinyatakan lebih mirip. Melanjut dari penerapan *glove* sebagai model untuk mengubah dataset menjadi sebuah *vector* maka yang dibandingkan adalah kemiripan *vector* dari dataset yang digunakan. Pada dataset yang digunakan pada penelitian ini yang diambil adalah kolom deskripsi dari data *user* dan data *job* yang kemudian diubah terlebih dahulu menjadi *vector* dan kemudian akan menghasilkan *similarity* dengan skor yang bermacam-macam dan yang diambil adalah yang memiliki hasil *similarity* terbesar atau mendekati 1.

4. Naïve Bayes

Pada tahap ini dilakukan perhitungan *naïve bayes* yang digunakan untuk menentukan probabilitas. *Naïve Bayes* itu sendiri merupakan sebuah metode probabilitas untuk klasifikasi. Metode ini dinilai dapat memberikan nilai akurasi yang kompetitif dan juga proses komputasi yang efisien. Dalam prakteknya ada beberapa model yang dapat digunakan diantaranya *gaussian*, *multinomial*, *bernouli*, dan *complement*. Setelah melakukan pengujian terhadap beberapa model di atas, didapatkan akurasi yang tinggi

pada model *gaussian*. Model *machine learning naïve bayes* menggunakan teori *Bayesian* (Perhitungan Probabilitas) untuk menentukan *outputnya*. Langkah-langkahnya adalah sebagai berikut :

1. Menentukan nilai probabilitas setiap fitur terhadap masing-masing kelas/target.
2. Hasil dari poin pertama dikalikan dengan probabilitas kemunculan masing-masing kelas.

$$P(F|Y)P(Y) = 0.13$$

$$P(F|Y)P(T) = 0.20$$

Berdasarkan pada data disamping nilai probabilitas kelas “tidak” lebih tinggi dibandingkan kelas “ya”. Sehingga, probabilitas “tidak yang paling tinggi. Begitu juga pada dataset yang digunakan dalam penelitian ini berjumlah 20 data, maka dengan probabilitas tertinggi akan dijadikan sebagai rekomendasi.

5. *Output*

Langkah akhir adalah dari hasil yang telah didapatkan pada penelitian ini dari dua metode tersebut dipilih kembali untuk ditentukan hasil rekomendasi pekerjaan apa yang paling tepat dan dideploy ke website menggunakan *flask*

5. *Evaluation*

Pada tahap terakhir ini, model yang sudah dibuat diuji dan dievaluasi keakuratan. Tahap ini mengukur sejauh mana model yang sudah dipilih memenuhi sasaran-sasaran bisnis dan bila demikian, sejauh manakah itu (apakah perlu lebih banyak model untuk dibuat dan diukur). Tahap terakhir yang dilakukan adalah pengujian yang dilakukan terhadap model untuk dievaluasi tingkat akurasinya. Bagian ini mengukur bagaimana model tersebut menangani suatu masalah, memberikan solusi terhadap masalah dan tujuan yang ingin dicapai.

4.2 Indikator Capaian Penelitian

Berdasarkan diagram alur metodologi penelitian diatas, terdapat indikator capaian sebagai berikut.

Tabel 4. 1 Indikator Capaian Penelitian

| No. | Tahapan | | Indikator capaian |
|-----|-------------------------------|---|--|
| 1. | <i>Business Understanding</i> | → | 1. <i>Flow chart</i> bisnis pada perusahaan xyz |
| 2. | <i>Data Understanding</i> | → | 2. Dataset <i>job</i> dan <i>user</i> yang belum dibersihkan yang terdiri dari <i>desc</i> pada dataset <i>job</i> dan <i>about</i> pada dataset <i>user</i> |
| 3. | <i>Data Preparation</i> | → | 3. Dataset yang sudah dibersihkan dengan kolom <i>about_clean</i> dan <i>desc_clean</i> . |
| 3. | <i>Modelling</i> | → | 4. Model Glove.txt dipilih untuk dapat memberikan hasil rekomendasi pekerjaan |
| 4. | Evaluasi | → | 5. Pengujian model |

BAB IV

HASIL IMPLEMENTASI

5.1 Implementasi

5.1.2. Lingkungan Implementasi

Lingkungan implementasi pada penelitian ini dilakukan oleh penulis pada rencana pembuatan sistem rekomendasi di salah satu *startup* untuk memenuhi kebutuhan pada proses bisnis yang terdapat pada bagian HRD. Adapun hasil penerapan sistem rekomendasi pekerjaan yang telah dibuat pada penelitian ini dijelaskan pada bagian hasil implementasi beserta pengujiannya.

5.2 Hasil

Bab ini menjawab hal-hal yang telah disampaikan pada bab sebelumnya tentang metodologi penelitian dari tahap *scraping*, *cleaning*, *cosine similarity*, *naïve bayes* dan evaluasi.

5.2.1. Hasil Scraping Data

Data diambil dari linkedin dan discraping menggunakan selenium. Pada tahap ini tahap scraping akan mengambil element yang terdapat pada halaman profil linkedin user kemudian akan disatukan menjadi kolom deskripsi seperti pada table di bawah ini :

| Linkedin profil | Deskripsi |
|---|---|
| https://www.linkedin.com/in/muhammad-fahmi-4429181a4/ | Hello, I'm Fahmi. Have an interest in working in an adaptable environment. Skilled in Analytical Thinking, Data Scientist and Python. Has experience focusing on Natural Language Processing and Data Mining. And I work as a Freelance and Instructor. |
| https://www.linkedin.com/in/rachma-adzima/ | Hello! I am Rachma Adzima Maulida, also known as Rachma. I extremely motivated to constantly develop my skills and grow professionally. I am a person who likes Web Programming, Mobile Programming and UI/UX Design. |
| https://www.linkedin.com/in/nanda-harsana-octavya/ | I am looking for opportunities in data analyst, data science and machine learning. Having experience in processing data using SQL, exploratory data analysis, data visualization, data processing and building machine learning models with Python. |

5.2.1. Pelabelan data

Langkah pertama yang dilakukan adalah melakukan pelabelan data, labelisasi data digunakan untuk melengkapi data yang belum didefinisikan. Pelabelan data dilakukan secara manual dengan mencari *job description* pada bidang tertentu.

| Deskripsi | Skill | Label |
|---|--|-----------------|
| Copywriters, or Marketing Writers, are responsible for producing engaging, clear text for different advertising channels such as websites, print ads and catalogs. Their duties include researching keywords, producing interesting written content and proofreading their work for accuracy and quality. | English Language Skills. A high level of English language skills are a must, An Eye For Detail, A Wide Vocabulary, Curiosity, The Ability To See Different Points of View, Research Skills, Great Listening Skills | Copy Writer |
| Content creators promote companies' products and services by developing copy for websites, social media, marketing materials, and other platforms. They create both written and verbal media for both private and public consumption | Content creation skills, Editing skills, Deep understanding of sales funnels, Planning skills, Research skills, SEO skills, Content promotion skills, Data analysis skills | Content Creator |
| UI UX designers create the user interface for an app, website, or other interactive media. Their work includes collaborating with product managers and engineers to gather requirements from users before designing ideas that can be communicated using storyboards. They also process flows or sitemaps. They'll then use their creativity while applying color palettes based on what best fits into a particular design conceptually. | Prototyping, wireframing, user flows, mockups. A huge part of the product development process is envisioning what a product will look like, Visual design and design software, User research and usability testing, Agile, Information architecture, Application development. We are looking for a UI/UX Designer to turn our software into easy-to-use products for our clients | UI/UX Design |

Data tersebut berdasarkan deskripsi pekerjaan yang diambil dari beberapa sumber seperti *linkedin*, *jobstreet*, dan platform *job search* lainnya.

5.2.2. Pre-processing

Data yang didapatkan umumnya berbentuk data mentah yang tidak terstruktur seperti terdapat link url, lowercase atau casefolding, punctuation dan lain-lain. Adanya data-data tersebut akan mempengaruhi kinerja saat melakukan modelling menggunakan *python*.

5.2.2.1. DROP NA

Drop NA bertujuan untuk membersihkan data yang kosong atau null secara otomatis menggunakan library *regex*.

5.2.2.2. Tokenization

Tokenization bertujuan untuk menguraikan teks menjadi perkata dan menghasilkan sebuah “token”. Sebelumnya huruf kapita diubah dulu menjadi huruf kecil.

| Teks asli | Hasil tokenization |
|--|--|
| Having experience in processing data using SQL | “having” “experience” “in” “processing” “data” “using” “sql” |

5.2.2.3. Remove Punctuation, Tags, Digits, Special Chars, Emoji & Other Symbol

Selanjutnya dataset dibersihkan kembali dari karakter-karakter seperti *url*, *hashtag*, *username*, tanda baca dan karakter alfanumerik terkecuali spasi seperti [, @, \, www, ^, https:, %, * dan lain-lain.

5.2.3. Naïve Bayes

Pada tahap ini dijelaskan tahap pembuatan model menggunakan *naïve bayes*. Tahap ini menentukan kualitas atau akurasi model yang akan digunakan. Adapun langkah-langkah yang dilakukan saat pembuatan model adalah sebagai berikut :

Melakukan pengujian terhaap 3 model

| Model | Test Size | Akurasi |
|----------------------|-----------|--------------------|
| MultinomialNB | 10% | 0.8157894736842105 |
| | 20% | 0.8157894736842105 |
| | 30% | 0.8370044052863436 |
| Gaussian | 10% | 0.868421052631579 |
| | 20% | 0.868421052631579 |
| | 30% | 0.8370044052863436 |
| Complement | 10% | 0.8421052631578947 |
| | 20% | 0.9078947368421053 |
| | 30% | 0.8678414096916299 |

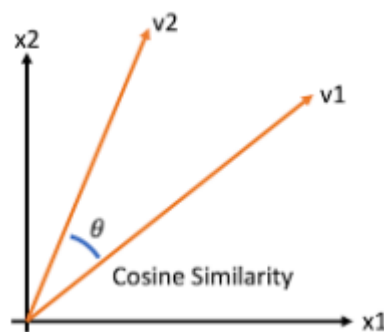
Dari pengujian yang dilakukan terhadap tiga model yang terdapat pada *naïve bayes* didapatkan hasil akurasi paling tinggi pada model *complement*. Model yang telah dibuat akan disave untuk kemudian digunakan saat proses deployment menggunakan *framework flask*.

5.2.4. Cosine Similarity

Langkah awal yang dilakukan adalah melakukan load dataset yang sebelumnya sudah dilabeli. Dataset yang digunakan berjumlah 20 yang sudah dilakukan pelabelan.

| | deskripsi | skill | label | label_num |
|---|---|---|-------------------|-----------|
| 0 | We are looking for Data Scientists who are int... | Probability & Statistics,Multivariate Calculus... | data scientist | 18 |
| 1 | About Rocket LawyerWe believe everyone deserve... | Python,Golang,Ruby,Perl,Scala,Java,SAS,R,MatLa... | data engineer | 17 |
| 2 | HR Development is concerned with the growth of... | Organisation,Time management,Communication,Con... | HRD | 10 |
| 3 | A Digital Marketing Manager is responsible for... | Search Engine Optimization (SEO) and Search En... | Digital Marketing | 8 |
| 4 | Perform haircuts and styleMust have NJ license... | Hairdresser skills are abilities and competenc... | Hair Stylist | 11 |
| 5 | A Product Manager is a professional who combin | Communication skills,Technical expertise,Busin... | Product Manager | 12 |

Data yang digunakan pada metode ini juga dilakukan tahap pre-processing sama halnya seperti yang dilakukan pada metode sebelumnya yang terdiri dari *casefolding*, *stopword*, dan *punctuation*. Setelah data di *cleaning*, dilakukan tahap pemodelan menggunakan model *glove*. Model *glove* merupakan proses *feature extraction* yang akan merubah data menjadi sebuah vector. Setelah model dibuat maka dilakukan tahap *cosine similarity*, Secara umum sistem yang dibangun menggunakan metode *Cosine Similairity* yaitu untuk menghitung *similarity* (tingkat kesamaan) antar dua buah objek misalnya $v1$ dan $v2$.



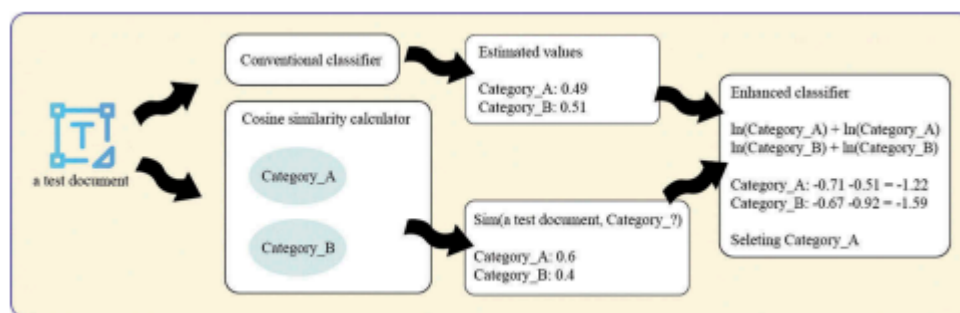
Dari *cosine similarity* dapat dicari tingkat kemiripan antara satu vector dan vector lain yang menghasilkan cosinus 0 derajat yang berarti 1. Maka semakin dekat kemiripan antar teks maka sudutnya akan semakin dekat dan suatu kata dikatakan mirip apabila mendekati 1. Dari proses *matching* menggunakan *cosine similarity* yang dites kepada salah satu *user* menghasilkan *similarity* sebagai berikut :

| | label_num | label | similarity |
|----|-----------|----------------------------|------------|
| 1 | 17 | data engineer | 0.908523 |
| 19 | 19 | mobile developer | 0.906299 |
| 14 | 7 | Data Business Analyst | 0.894999 |
| 6 | 13 | Quality Assurance | 0.892800 |
| 13 | 2 | Brand Ambassador | 0.891271 |
| 3 | 8 | Digital Marketing | 0.890230 |
| 18 | 15 | Sales Manager | 0.890215 |
| 20 | 20 | web developer / programmer | 0.889273 |
| 0 | 18 | data scientist | 0.885046 |
| 11 | 6 | Copywriter | 0.882503 |
| 17 | 4 | Business Development | 0.880910 |
| 15 | 14 | Quality Control | 0.875438 |

Hasil *similarity* pada label *data engineer* mendapatkan hasil lebih besar dan paling mendekati 1, maka hasil tersebut dianggap label yang paling *similar*.

5.2.5. Combine Naïve bayes dan Cosine similarity

Hasil yang didapat dari masing-masing metode akan dikombinasikan, yaitu metode *naïve bayes* dan metode *cosine similarity*. Hasil dari *naïve bayes* merupakan angka probabilitas dan *cosine similarity* adalah *similarity*.



Untuk hasil rekomendasi tertinggi akan diurutkan untuk kemudian dimasukkan ke dalam formula untuk mendapatkan kalkulasi persentasi antara kedua hasil tersebut seperti berikut ini :

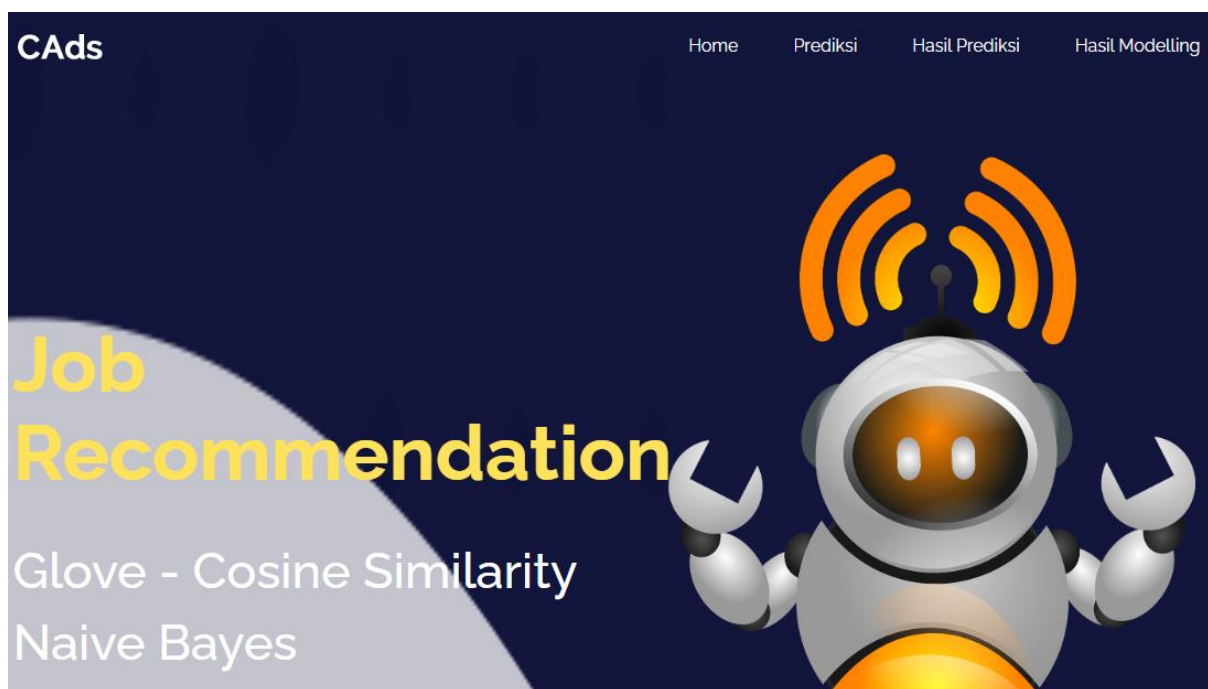
```
combine['kalkulasi_persen'] = (combine['probability_naivebayes'] + combine['similarity_cosine']) / 2
```

| | recommendation | probability_naivebayes | label | similarity_cosine | kalkulasi_persen |
|----|-----------------------|------------------------|-------|-------------------|------------------|
| 18 | Data scientist | 75.706147 | 18 | 89.021477 | 82.363812 |
| 7 | Data business analyst | 62.438504 | 7 | 85.410370 | 73.924437 |
| 17 | Data engineer | 56.981438 | 17 | 88.091001 | 72.536219 |
| 19 | Mobile developer | 49.507737 | 19 | 90.629870 | 70.068803 |
| 14 | Quality control | 49.196541 | 14 | 89.499854 | 69.348198 |
| 6 | Copywriter | 48.670219 | 6 | 89.279967 | 68.975093 |
| 20 | Web developer | 48.952235 | 20 | 88.927331 | 68.939783 |

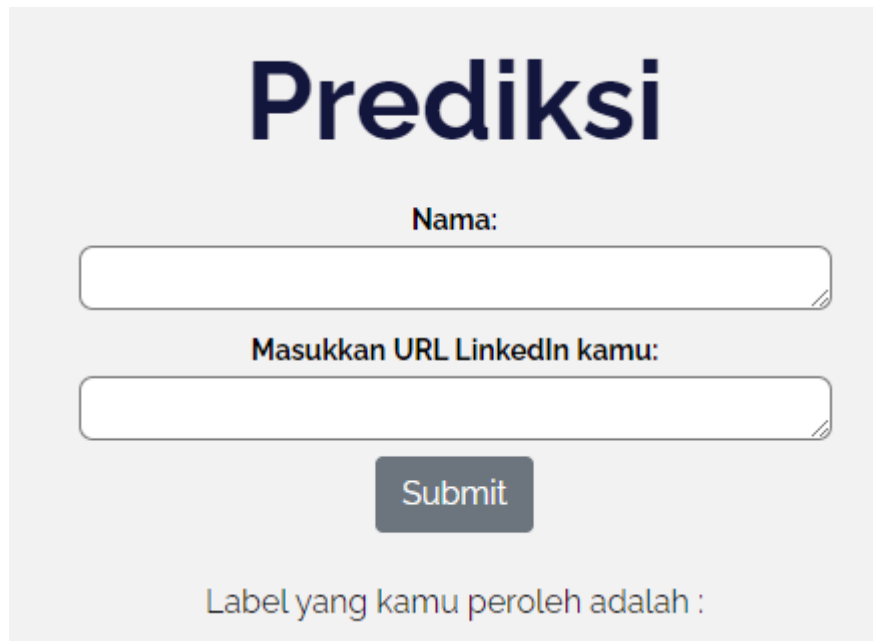
Dari hasil tersebut dapat diketahui bahwa user 1 direkomendasikan dengan label 18 yaitu *data scientist* dengan hasil probability 75% dan *similarity* 89% yang dikalkulasikan menjadi 82%.

5.2.5. Deployment

Untuk melakukan deployment model akan menggunakan flask sebagai framework yang sudah banyak digunakan dalam pembuatan deployment machine learning.



Tampilan Home



Prediksi

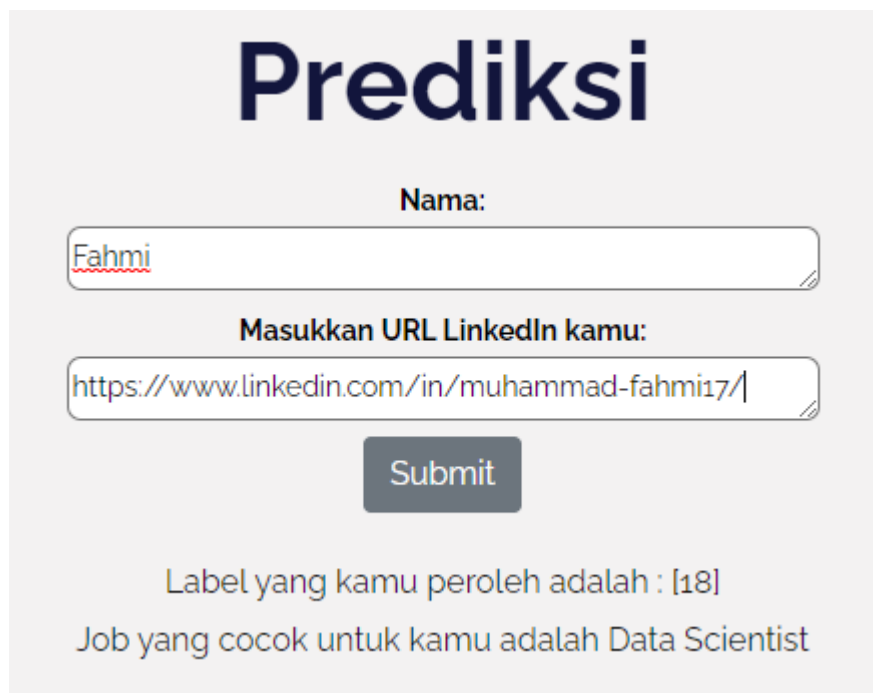
Nama:

Masukkan URL LinkedIn kamu:

Submit

Label yang kamu peroleh adalah :

Tampilan Halaman Prediksi



Prediksi

Nama:

Masukkan URL LinkedIn kamu:

Submit

Label yang kamu peroleh adalah : [18]
Job yang cocok untuk kamu adalah Data Scientist

Tampilan halaman hasil prediksi

BAB VI

PENUTUP

1.1 Kesimpulan dan Saran

Berdasarkan hasil analisis pada laporan internship II dapat disimpulkan bahwa :

1. Sistem rekomendasi menghasilkan rangkaian pekerjaan yang dianggap cocok dengan user.
2. Sistem rekomendasi memberikan rekomendasi pekerjaan berdasarkan profil *user* di linkedin.

Berdasarkan hasil analisis pada laporan intermship II disarankan untuk :

1. Melakukan perbaikan pada hasil *similarity*.
2. Melakukan perbandingan metode *supervised* dan *unsupervised*.
3. Akurasi hasil *cosine similarity* harus dapat diukur.

DAFTAR PUSTAKA

- Alkaff, M., Khatimi, H., & Eriadi, A. (2020). Sistem Rekomendasi Buku pada Perpustakaan Daerah Provinsi Kalimantan Selatan Menggunakan Metode Content-Based Filtering. *MATRIK : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 20(1), 193–202. <https://doi.org/10.30812/matrik.v20i1.617>
- Asfi, M., & Fitrianiingsih, N. (2020). Implementasi Algoritma Naive Bayes Classifier sebagai Sistem Rekomendasi Pembimbing Skripsi. *InfoTekJar : Jurnal Nasional Informatika Dan Teknologi Jaringan*, 5(1), 44–50. <https://doi.org/10.30743/infotekjar.v5i1.2536>
- Badriyah, T., Fernando, R., & Syarif, I. (2018). Sistem Rekomendasi Content Based Filtering Menggunakan Algoritma Apriori. *Konferensi Nasional Sistem Informasi*, 1(1), 554–559.
- Bag, S., Ghadge, A., & Tiwari, M. K. (2019). An integrated recommender system for improved accuracy and aggregate diversity. *Computers and Industrial Engineering*, 130(April), 187–197. <https://doi.org/10.1016/j.cie.2019.02.028>
- Budiman, M. A., & Mastrika Giri, G. A. V. (2020). Song Recommendations Based on Artists with Cosine Similarity Algorithms and K-Nearest Neighbor. *JELIKU (Jurnal Elektronik Ilmu Komputer Udayana)*, 8(4), 367. <https://doi.org/10.24843/jlk.2020.v08.i04.p01>
- Das, D., Sahoo, L., & Datta, S. (2017). A Survey on Recommendation System. *International Journal of Computer Applications*, 160(7), 6–10. <https://doi.org/10.5120/ijca2017913081>
- Deldjoo, Y., Dacrema, M. F., Constantin, M. G., Eghbal-zadeh, H., Cereda, S., Schedl, M., Ionescu, B., & Cremonesi, P. (2019). Movie genome: alleviating new item cold start in movie recommendation. *User Modeling and User-Adapted Interaction*, 29(2), 291–343. <https://doi.org/10.1007/s11257-019-09221-y>
- Desai, V., Bahl, D., Vibhandik, S., & Fatma, I. (2017). Implementation of an Automated Job Recommendation System Based on Candidate Profiles. *International Research Journal of Engineering and Technology (IRJET)*, 4(5), 1018–1021. <https://www.irjet.net/archives/V4/i5/IRJET-V4I5343.pdf>
- Ferio, G., Intan, R., & Rostianingsih, S. (2019). Sistem Rekomendasi Mata Kuliah Pilihan Menggunakan Metode User Based Collaborative Filtering Berbasis Algoritma Adjusted Cosine Similarity. *Jurnal Infra*, 7(1), 1–7.
- Fu, M., Qu, H., Yi, Z., Lu, L., & Liu, Y. (2019). A Novel Deep Learning-Based Collaborative Filtering Model for Recommendation System. *IEEE Transactions on Cybernetics*, 49(3), 1084–1096. <https://doi.org/10.1109/TCYB.2018.2795041>
- Gautam, A., & Jerripothula, K. R. (2020). SGG: Spinbot, Grammarly and GloVe based Fake News Detection. *Proceedings - 2020 IEEE 6th International Conference on Multimedia Big Data, BigMM 2020*, 174–182. <https://doi.org/10.1109/BigMM50055.2020.00033>
- Govindarajan, M. (2020). *Introduction to Data Science*. 2012, 1–18. <https://doi.org/10.4018/978-1-7998-3053-5.ch001>
- Grashinta, A., Istiqomah, A. P., & Wiroko, E. P. (2018). Pengaruh Future Time Perspective Terhadap Kematangan Karir Pada Mahasiswa. *Jurnal Psikologi Pendidikan Dan Konseling: Jurnal Kajian Psikologi Pendidikan Dan Bimbingan Konseling*, 4(1), 25. <https://doi.org/10.26858/jpkk.v4i1.4981>

- Husein, A. M., & Harahap, M. (2021). Pendekatan Data Science untuk Menemukan Churn Pelanggan pada Sector Perbankan dengan Machine Learning. *Data Sciences Indonesia (DSI)*, 1(1), 8–13. <https://doi.org/10.47709/dsi.v1i1.1169>
- Hwangbo, H., Kim, Y. S., & Cha, K. J. (2018). Recommendation system development for fashion retail e-commerce. *Electronic Commerce Research and Applications*, 28, 94–101. <https://doi.org/10.1016/j.elerap.2018.01.012>
- Ibrahim, M., Gauch, S., Gerth, T., & Cox, B. (2021). *WOVe: Incorporating Word Order in GloVe Word Embeddings*. <https://arxiv.org/abs/2105.08597>
- Ibrahim, M., Gauch, S., Salman, O., & Alqahatani, M. (2020). Enriching consumer health vocabulary using enhanced GloVe word embedding. *ArXiv Preprint ArXiv:2004.00150*.
- Intan Aprilia, R., & Fachrurrozi, M. (2016). *menggunakan Metode Collaborative Filtering dan Naive Bayes*. 2(1), 343–347. <http://ars.ilkom.unsri.ac.id343>
- Jepriana, I. W., & Wardoyo, R. (2018). Algoritme Genetika untuk Mengurangi Galat Prediksi Metode Item-based Collaborative Filtering. *Bimipa*, 25(2), 165–174.
- Jiang, J.-Y., Zhang, M., Li, C., Bendersky, M., Golbandi, N., & Najork, M. (2019). Semantic text matching for long-form documents. *The World Wide Web Conference*, 795–806.
- Jiang, L., Cheng, Y., Yang, L., Li, J., Yan, H., & Wang, X. (2019). A trust-based collaborative filtering algorithm for E-commerce recommendation system. In *Journal of Ambient Intelligence and Humanized Computing* (Vol. 10, Issue 8, pp. 3023–3034). <https://doi.org/10.1007/s12652-018-0928-7>
- Kenthapadi, K., Le, B., & Venkataraman, G. (2017). Personalized job recommendation system at LinkedIn: Practical challenges and lessons learned. *RecSys 2017 - Proceedings of the 11th ACM Conference on Recommender Systems*, 346–347. <https://doi.org/10.1145/3109859.3109921>
- Lakmal, D., Ranathunga, S., Peramuna, S., & Herath, I. (2020). Word embedding evaluation for Sinhala. *LREC 2020 - 12th International Conference on Language Resources and Evaluation, Conference Proceedings, May*, 1874–1881.
- Ma, W., Shi, J., & Zhao, R. (2017). Normalizing item-based collaborative filter using context-aware scaled baseline predictor. *Mathematical Problems in Engineering*, 2017. <https://doi.org/10.1155/2017/6562371>
- Martinez-Plumed, F., Contreras-Ochando, L., Ferri, C., Hernandez-Orallo, J., Kull, M., Lachiche, N., Ramirez-Quintana, M. J., & Flach, P. (2021). CRISP-DM Twenty Years Later: From Data Mining Processes to Data Science Trajectories. *IEEE Transactions on Knowledge and Data Engineering*, 33(8), 3048–3061. <https://doi.org/10.1109/TKDE.2019.2962680>
- Mohammed, S. M., Jacksi, K., & Zeebaree, S. R. M. (2021). A state-of-the-art survey on semantic similarity for document clustering using GloVe and density-based algorithms. *Indonesian Journal of Electrical Engineering and Computer Science*, 22(1), 552–562. <https://doi.org/10.11591/ijeecs.v22.i1.pp552-562>
- Mustika, D. (2017). Disusun Oleh : Disusun Oleh : *Pelaksanaan Pekerjaan Galian Diversion Tunnel Dengan Metode Blasting Pada Proyek Pembangunan Bendungan Leuwikeris Paket 3, Kabupaten Ciamis Dan Kabupaten Tasikmalaya Jawa Barat*,

1(11150331000034), 1–147.

- Nastiti, P. (2019). Penerapan Metode Content Based Filtering Dalam Implementasi Sistem Rekomendasi Tanaman Pangan. *Teknika*, 8(1), 1–10. <https://doi.org/10.34148/teknika.v8i1.139>
- Nasution, M. K. M. (2021). Industry 4.0: Data science perspective. *IOP Conference Series: Materials Science and Engineering*, 1122(1), 012037. <https://doi.org/10.1088/1757-899x/1122/1/012037>
- Nugroho. (2020). Sistem Rekomendasi Produk Ukm Di Kota Bandung Menggunakan Algoritma Collaborative Filtering. *Jurnal Riset Sistem Informasi Dan Teknologi Informasi (JURSISTEKNI)*, 2(3), 23–31.
- Nurdin, A., Aji, B. A. S., Bustamin, A., & Abidin, Z. (2020). Perbandingan Kinerja Word Embedding Word2Vec, Glove, Dan Fasttext Pada Klasifikasi Teks. *Jurnal Tekno Kompak*, 14(2), 74–79.
- Portugal, I., Alencar, P., & Cowan, D. (2018). The use of machine learning algorithms in recommender systems: A systematic review. *Expert Systems with Applications*, 97, 205–227. <https://doi.org/10.1016/j.eswa.2017.12.020>
- Prasetya, W. S. (2017). Perancangan model basis data relasional dengan metode Database Life Cycle. *Seminar Nasional Informatika (SNIIf)*, 1(1), 91–98.
- Putri, D. C. G., Leu, J. S., & Seda, P. (2020). Design of an unsupervised machine learning-based movie recommender system. *Symmetry*, 12(2), 1–27. <https://doi.org/10.3390/sym12020185>
- Putri, M. W., Muchayan, A., & Kamisutara, M. (2020). Sistem Rekomendasi Produk Pena Eksklusif Menggunakan Metode Content-Based Filtering dan TF-IDF. *JOINTECS (Journal of Information Technology and Computer Science)*, 5(3), 229. <https://doi.org/10.31328/jointecs.v5i3.1563>
- Rahman, M. Z., Sari, Y. A., & Yudistira, N. (2021). Analisis Sentimen Tweet COVID-19 menggunakan Word Embedding dan Metode Long Short-Term Memory (LSTM). 5(11), 5120–5127.
- Rahmawati, S., Nurjanah, D., & Rismala, R. (2018). Analisis dan Implementasi pendekatan Hybrid untuk Sistem Rekomendasi Pekerjaan dengan Metode Knowledge Based dan Collaborative Filtering. *Indonesian Journal on Computing (Indo-JC)*, 3(2), 11. <https://doi.org/10.21108/indojc.2018.3.2.210>
- Reddy, S., Nalluri, S., Kuniseti, S., Ashok, S., & Venkatesh, B. (2019). Content-based movie recommendation system using genre correlation. *Smart Innovation, Systems and Technologies*, 105, 391–397. https://doi.org/10.1007/978-981-13-1927-3_42
- Retnoningsih, E., & Pramudita, R. (2020). Mengenal Machine Learning Dengan Teknik Supervised Dan Unsupervised Learning Menggunakan Python. *Bina Insani Ict Journal*, 7(2), 156. <https://doi.org/10.51211/biiict.v7i2.1422>
- Ruehle, F. (2020). Data science applications to string theory. *Physics Reports*, 839, 1–117. <https://doi.org/10.1016/j.physrep.2019.09.005>
- Sanchez-Pinto, L. N., Luo, Y., & Churpek, M. M. (2018). Big Data and Data Science in Critical

- Care. *Chest*, 154(5), 1239–1248. <https://doi.org/10.1016/j.chest.2018.04.037>
- Shalaby, W., Alaila, B. E., Korayem, M., Pournajaf, L., Aljadda, K., Quinn, S., & Zadrozny, W. (2017). Help me find a job: A graph-based approach for job recommendation at scale. *Proceedings - 2017 IEEE International Conference on Big Data, Big Data 2017, 2018-Janua*, 1544–1553. <https://doi.org/10.1109/BigData.2017.8258088>
- Simbolon, J., & Kasih, J. (2019). *Collaborative Filtering Dan Perancangan Website Gigidiyor*. 1, 227–236.
- Singh, R. H., Maurya, S., Tripathi, T., Narula, T., & Srivastav, G. (2020). Movie Recommendation System using Cosine Similarity and KNN. *International Journal of Engineering and Advanced Technology*, 9(5), 556–559. <https://doi.org/10.35940/ijeat.e9666.069520>
- Suka Parwita, W. G., Prami Swari, M. H., & Welda, W. (2018). Perancangan Sistem Rekomendasi Dokumen Dengan Pendekatan Content-Based Filtering. *Computer Engineering, Science and System Journal*, 3(1), 65. <https://doi.org/10.24114/cess.v3i1.7855>
- Wang, B., Wang, A., Chen, F., Wang, Y., & Kuo, C. C. J. (2019). Evaluating word embedding models: Methods and experimental results. *APSIPA Transactions on Signal and Information Processing*, 8. <https://doi.org/10.1017/ATSIP.2019.12>
- Webb, G. I. (2020). Encyclopedia of Machine Learning and Data Science. *Encyclopedia of Machine Learning and Data Science*, April. <https://doi.org/10.1007/978-1-4899-7502-7>
- Wijaya, A. E., & Alfian, D. (2018). Sistem Rekomendasi Laptop Menggunakan Collaborative Filtering Dan Content-Based Filtering. *Jurnal Computech & Bisnis*, 12(1), 11–27.
- Yang, S., Korayem, M., AlJadda, K., Grainger, T., & Natarajan, S. (2017). Combining content-based and collaborative filtering for job recommendation system: A cost-sensitive Statistical Relational Learning approach. *Knowledge-Based Systems*, 136, 37–45. <https://doi.org/10.1016/j.knosys.2017.08.017>
- Yasni, L., Subroto, I. M. I., & Haviana, S. F. C. (2018). Implementasi Cosine Similarity Matching Dalam Penentuan Dosen Pembimbing Tugas Akhir. *Transmisi*, 20(1), 22. <https://doi.org/10.14710/transmisi.20.1.22-28>
- Zhang, J., Wang, Y., Yuan, Z., & Jin, Q. (2020). Personalized real-time movie recommendation system: Practical prototype and evaluation. *Tsinghua Science and Technology*, 25(2), 180–191. <https://doi.org/10.26599/TST.2018.9010118>
- Zhang, Z., Kudo, Y., & Murai, T. (2017). Neighbor selection for user-based collaborative filtering using covering-based rough sets. *Annals of Operations Research*, 256(2), 359–374. <https://doi.org/10.1007/s10479-016-2367-1>

LAMPIRAN-LAMPIRAN

Lampiran 1. Bukti Kartu Bimbingan (KAMBING)

Lampiran 1 Lembar Kartu Bimbingan (KAMBING)



POLITEKNIK POS INDONESIA
PROGRAM STUDI D4 TEKNIK INFORMATIKA
JL. SARIASIH NO. 54 BANDUNG 40151
Telp. 022-2009562, 2009570
Fax. 022-2009568

FORMULIR KEGIATAN

INTERNSHIP II

TA. 2021/2022

Nama : Muhammad Dzihan Al-Banna
Npm : 1174095
Judul : Iiteung Penerapan Sistem Rekomendasi Pada Job Recommendation Berdasarkan Profil LinkedIn Menggunakan Cosine Similarity
Pembimbing : Roni Habibi, S.Kom., M.T.,SFPC



| Pertemuan | Tanggal | Sudah Dikerjakan | Pekerjaan Selanjutnya | Nilai |
|-----------|------------|---|---|-------|
| 1 | 22-04-2022 | pengajuan dan pembahasan judul internship 2 | pembahasan mengenai proposal internship 2 | 75 |
| 2 | 10-05-2022 | presentasi proposal terkait metode dan tinjauan pustaka | memperbaiki proposal mengenai tinjauan pustaka serta metode penelitian | 87 |
| 3 | 21-05-2022 | pembahasan metode penelitian dan bab 1 | pelajari metode ken peffers | 85 |
| 4 | 27-05-2022 | tambah metode penelitian | melanjutkan ke bab selanjutnya | 87 |
| 5 | 31-05-2022 | bab 1 laporan internship 2 | melanjutkan ke bab selanjutnya dan menambahkan referensi | 85 |
| 6 | 09-06-2022 | bab 3 dan 4 laporan internship 2 | revisi bab 3 dan 4 serta menambahkan jurnal | 85 |
| 7 | 16-06-2022 | bab 3 dan 4 laporan internship 2 | menambahkan analisis serta perbaikan pada bab 3 dan 4 serta menambahkan referensi | 87 |
| 8 | 23-06-2022 | terkait bab 4 | melanjutkan ke bab selanjutnya | 90 |
| 9 | 29-06-2022 | diskusi mengenai metode penelitian | melanjutkan ke laporan bab 5 simulasi data | 90 |
| 10 | 05-07-2022 | presentasi mengenai penelitian | melanjutkan ke draft jurnal | 90 |
| | | | Rata-Rata: | 86.10 |

Bandung, 07 Juli 2022

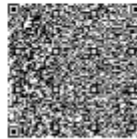
Pembimbing,

| | | | | |
|----|------------|--|---|-------|
| 3 | 21-05-2022 | pembahasan metode penelitian dan bab 1 | pelajari metode ken pelfers | 85 |
| 4 | 27-05-2022 | tambah metode penelitian | melanjutkan ke bab selanjutnya | 87 |
| 5 | 31-05-2022 | bab 1 laporan internship 2 | melanjutkan ke bab selanjutnya dan menambahkan referensi | 85 |
| 6 | 09-06-2022 | bab 3 dan 4 laporan internship 2 | revisi bab 3 dan 4 serta menambahkan jurnal | 85 |
| 7 | 16-06-2022 | bab 3 dan 4 laporan internship 2 | menambahkan analisis serta perbaikan pada bab 3 dan 4 serta menambahkan referensi | 87 |
| 8 | 23-06-2022 | terkait bab 4 | melanjutkan ke bab selanjutnya | 90 |
| 9 | 29-06-2022 | diskusi mengenai metode penelitian | melanjutkan ke laporan bab 5 simulasi data | 90 |
| 10 | 05-07-2022 | presentasi mengenai penelitian | melanjutkan ke draft jurnal | 90 |
| | | | Rata-Rata: | 86.10 |

Bandung, 07 Juli 2022

Pembimbing,

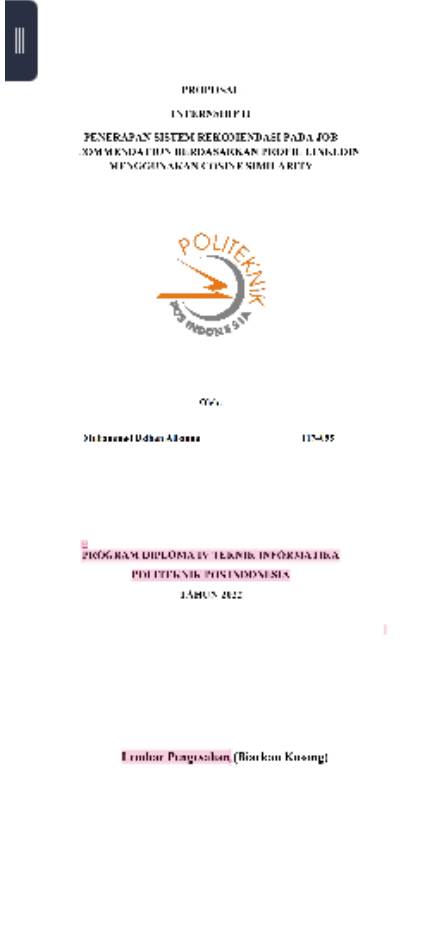
37b6c8bf-667-4c74-ba77-167bca3a212



Roni Habibi, S.Kom., M.T., SFPC
NIDN. 0423127804

Lampiran 2. Bukti Pengecekan Plagiarisme Online

Lampiran 2 Lembar Bukti Pengecekan Plagiarisme Online



20

Match Overview

20%

<

>

| | | |
|---|--|------|
| 1 | repository.unsil.ac.id Internet Source | 2% > |
| 2 | www.coursehero.com Internet Source | 1% > |
| 3 | Dspace.Uji.Ac.Id Internet Source | 1% > |
| 4 | Submitted to SDM Univ... Student Paper | 1% > |
| 5 | ejournal.undip.ac.id Internet Source | 1% > |
| 6 | journal.universitasbumi... Internet Source | 1% > |
| 7 | es.scribd.com Internet Source | 1% > |

Lampiran 3. Surat Pernyataan

Lampiran 3 Surat Pernyataan

SURAT PERNYATAAN TIDAK MELAKUKAN PLAGIARISME

Yang bertanda tangan di bawah ini :

Nama : Muhammad Dzihan Albanna

NPM : 1174095

Program Studi : DIV Teknik Informatika

Judul : Penerapan Sistem Rekomendasi pada Job Recommendation Berdasarkan Profile LinkedIn Menggunakan Cosine Similarity

Menyatakan bahwa :

1. Proyek Pemrograman Aplikasi (INTERNSHIP II) saya ini adalah asli dan belum pernah diajukan untuk memenuhi kelulusan matakuliah Internship I pada Program Studi DIV Teknik Informatika baik di Politeknik Pos Indonesia maupun di Perguruan Tinggi lainnya.
2. Proyek Pemrograman Aplikasi (INTERNSHIP II) ini adalah murni gagasan, rumusan, dan penelitian saya sendiri tanpa bantuan pihak lain, kecuali arahan pembimbing.
3. Dalam Proyek Pemrograman Aplikasi (INTERNSHIP II) ini tidak terdapat karya atau pendapat yang telah ditulis atau dipublikasikan orang lain, kecuali secara tertulis dengan jelas dicantumkan sebagai acuan dalam naskah dengan disebutkan nama pengarang dan dicantumkan dalam daftar pustaka.
4. Pernyataan ini saya buat dengan sesungguhnya dan apabila di kemudian hari terdapat penyimpangan-penyimpangan dan ketidakbenaran dalam pernyataan ini, maka saya bersedia menerima sanksi akademik berupa pencabutan gelar yang telah diperoleh karena karya ini, serta sanksi lainnya sesuai dengan norma yang berlaku diperguruan tinggi lain.

Bandung, 14 Juli 2022

Yang Membuat Pernyataan,

Muhammad Dzihan Albanna

NPM. 117.4.095

Lampiran 4. Draft Jurnal

Lampiran 4 Draft Jurnal

Penerapan Sistem Rekomendasi Pada Job Recommendation Berdasarkan Profil LinkedIn Menggunakan Cosine Similarity

Implementation of a Recommendation System on Job Recommendations Based on LinkedIn Profiles Using Cosine Similarity

Roni Habibi¹, Muhammad Dzihan Al-Banna²

^{1,2}Politeknik Pos Indonesia, Bandung, Indonesia

³Contoh: Program Studi D4 Teknik Informatika, Politeknik Pos Indonesia, Bandung, Indonesia

*E-mail: muhammaddzihan55@gmail.com

Abstrak

Data adalah kumpulan informasi yang diperoleh dari beberapa sumber untuk disajikan dalam dunia bisnis. Ilmu data adalah teknologi untuk memecahkan masalah analitis yang kompleks. Dalam ilmu data juga terdapat analisis prediktif terhadap data yang akan disaring dan ditemukan data yang menghasilkan data yang akurat sesuai dengan data yang sebenarnya. Data juga banyak digunakan dalam kehidupan salah satunya digunakan dalam aplikasi linkedin dimana dengan menggunakan aplikasi dapat menemukan orang yang dikenali. Fitur ini menyarankan daftar dari orang-orang yang mungkin dikenal, yang mirip dengan data personal satu sama lain seperti lokasi saat ini atau mungkin lokasi masa lalu, keahlian, kelompok, halaman yang disukai, dan sebagainya. Rekomendasi ini khusus untuk personal dan berbeda dari pengguna ke pengguna. Untuk mendukung hal tersebut penulis akan membuat sistem rekomendasi pekerjaan menggunakan bahasa pemrograman python serta didukung oleh Recommendation System. Proses untuk menghasilkan output berupa daftar pekerjaan yang direkomendasikan dimulai dari pengumpulan data job dan user linkedin. Pada sistem rekomendasi penulis menggunakan content-based filtering atau content based recommendation. Content Based Recommendation memanfaatkan informasi beberapa item/ data untuk digunakan sebagai referensi yang terkait dengan informasi yang digunakan sebelumnya.

Kata kunci: Sistem Informasi; Rekomendasi; Content based recommendation.

Abstract

Data is a collection of information obtained from several sources to be presented in the business world. Data science is a technology for solving complex analytical problems. In data science there is also predictive analysis of the data to be filtered and found data that produces accurate data according to the actual data. Data is also widely used in life, one of which is used in the linkedin application where by using the application you can find recognized people. This feature suggests a list of people you might know, who are similar to you based on your friends, friends of friends in your circle of friends, current or possibly past locations, skills, groups, liked pages, and so on. These recommendations are specific to you and vary from user to user. To support this, the author will create a job recommendation system using the python programming language and supported by the Recommendation System. The process to produce output in the form of a list of recommended jobs starts from collecting job data and linkedin users. In the recommendation system, the author uses content-based filtering or content-based recommendation. Content Based Recommendation utilizes information on several items/data to be used as references related to previously used information.

Keywords: Information System; Recommendation; Content based recommendations..

Naskah diterima 2 Jan. 2022; direvisi 1 Feb. 2022; dipublikasikan 1 Apr. 2022.
JAMIKA is licensed under a Creative Commons Attribution-ShareAlike 4.0 International License.



1. PENDAHULUAN

1.1. Latar Belakang

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 dimasukkan 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 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 mengungguli contoh lainnya pada word analogy, word similarity, & named entity recognition [6]. Setelah itu data akan diolah menggunakan metode cosine similarity yang biasa diterapkan secara luas di pencarian informasi dan studi terkait[7].

1.2. Identifikasi Masalah

Adapun identifikasi masalah pada penelitian ini sebagai berikut:

- a. Bagaimana cara mempercepat jobseeker dalam memfilter lowongan pekerjaan?
- b. Bagaimana memberikan rekomendasi pekerjaan yang sesuai bagi jobseeker?
- c. Bagaimana menampilkan hasil dari rekomendasi pekerjaan yang mudah dipahami?

1.3. Tujuan dan Manfaat

Adapun tujuan pada penelitian ini sebagai berikut:

- a. Memfilter lowongan pekerjaan yang dinilai sesuai dengan jobseeker
- b. Memberikan rekomendasi pekerjaan yang sesuai bagi jobseeker berdasarkan profil.
- c. Website menggunakan flask sebagai framework untuk memvisualisasikan hasil rekomendasi.

Adapun manfaat pada penelitian ini sebagai berikut:

- a. Dapat memberikan rekomendasi pekerjaan yang relevan berdasarkan profil linkedin.
- b. Dapat memberikan peringkat kecocokan antara skill dengan job desc.

1.4. Ruang Lingkup

Adapun ruang lingkup pada penelitian ini sebagai berikut:

- a. Data yang digunakan untuk pengujian adalah data dari user linkedin yang sudah dikumpulkan melalui google form.
- b. Pengujian hanya dilakukan pada data yang didapatkan saat penyebaran google form.
- c. Penelitian hanya dilakukan pada lingkungan tempat peneliti melakukan Internship II.

2. LANDASAN TEORI

2.1.1. Tinjauan Pustaka

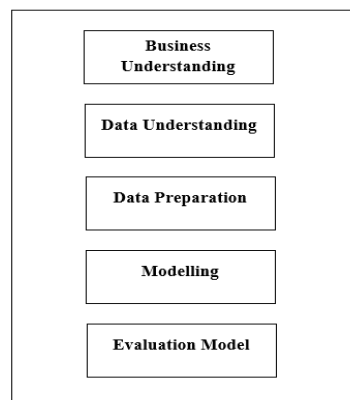
Sistem rekomendasi bertujuan untuk memberikan sebuah rekomendasi yang memancing perhatian user misalnya rekomendasi barang, rekomendasi teman, rekomendasi film, musik dan lain-lain[7]. Sistem rekomendasi menggunakan data yang diolah menjadi sebuah informasi untuk membuat suatu rekomendasi bagi user. 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[11].

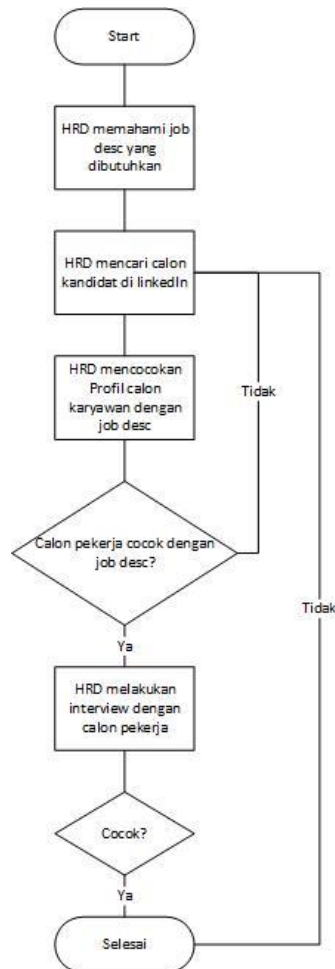
3. METODE PENELITIAN



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 pada 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 perbandingan. 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

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.

JOB RECOMMENDATION

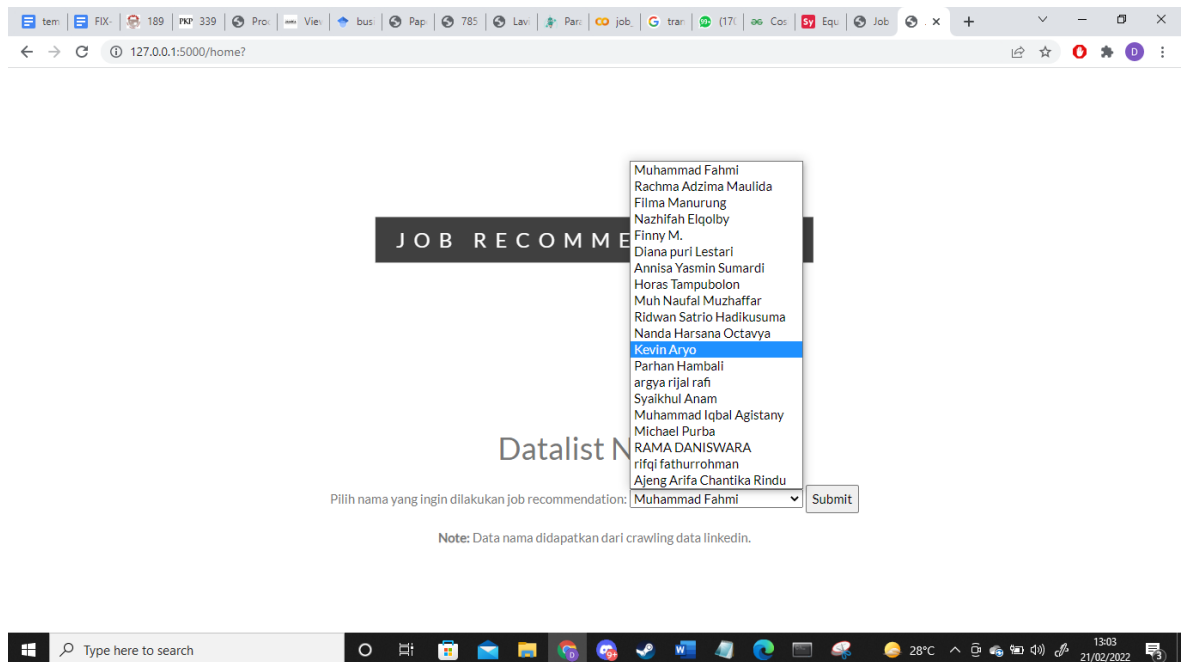
Datalist Name

Pilih nama yang ingin dilakukan job recommendation:

Note: Data nama didapatkan dari crawling data linkedin.

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 4.

Hasil Job Recommendation

Hasil 10 similarity teratas

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

Hasil 10 similarity terbawah

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

Gambar 5. Hasil rekomendasi pekerjaan

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

5. KESIMPULAN

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 masih rendah dikarenakan dataset job yang reduksi sehingga terdapat rekomendasi pekerjaan yang sama pada hasil akhir seperti data scientist machine learning security, data scientist, data scientist – cyber incident analysis

UCAPAN TERIMA KASIH

Terimakasih saya ucapkan kepada pihak-pihak yang telah membantu dalam pembuatan jurnal ini.

DAFTAR PUSTAKA

- [1] Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big data*, 1(1), 51-59.
- [2] Das, D., Sahoo, L., & Datta, S. (2017). A survey on recommendation system. *International Journal of Computer Applications*, 160.
- [3] Prasetya, C. S. D. (2017). Sistem Rekomendasi Pada E-Commerce Menggunakan K-Nearest Neighbor. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK)* p-ISSN, 2355, 7699.
- [4] Aditya, M. A., Mulyana, R. D., Eka, I. P., & Widiyanto, S. R. (2020, February). Penggabungan Teknologi Untuk Analisa Data Berbasis Data Science. In *Seminar Nasional Teknologi Komputer & Sains (SAINTEKS)* (Vol. 1, No. 1, pp. 51-56)

- [3] Das, D., Sahoo, L., & Datta, S. (2017). A survey on recommendation system. *International Journal of Computer Applications*, 160(7).
- [4] Grashinta, A., Istiqomah, A. P., & Wiroko, E. P. (2018). Pengaruh Future Time Perspective Terhadap Kematangan Karir Pada Mahasiswa. *Jurnal Psikologi Pendidikan Dan Konseling: Jurnal Kajian Psikologi Pendidikan Dan Bimbingan Konseling*, 4(1), 25. <https://doi.org/10.26858/jpkk.v4i1.4981>
- [5] Wijaya, A., & Alfian, D. (2018). Sistem Rekomendasi Laptop Menggunakan Collaborative Filtering Dan Content-Based Filtering. *Jurnal Computech & Bisnis*, 12(1), 11-27.
- [6] Indrapurasih, R. D., Bijaksana, M. A., & Sardi, I. L. (2018). Implementasi dan Analisis Kesamaan Semantik Antar Kata Bahasa Indonesia Menggunakan Metode GloVe. *eProceedings of Engineering*, 5(3).
- [7] Rahutomo, F., Kitasuka, T., & Aritsugi, M. (2012, October). Semantic cosine similarity. In *The 7th International Student Conference on Advanced Science and Technology ICAST* (Vol. 4, No. 1).
- [8] Badriyah, T., Fernando, R., & Syarif, I. (2018). Sistem Rekomendasi Content Based Filtering Menggunakan Algoritma Apriori. *Konferensi Nasional Sistem Informasi (KNSI) 2018*.
- [9] Nastiti, P. (2019). Penerapan Metode Content Based Filtering dalam Implementasi Sistem Rekomendasi Tanaman Pangan. *Teknika*, 8(1), 1-10.
- [10] Wijaya, A., & Alfian, D. (2018). Sistem Rekomendasi Laptop Menggunakan Collaborative Filtering Dan Content-Based Filtering. *Jurnal Computech & Bisnis*, 12(1), 11-27.
- [11] Mikolov, T., Grave, E., Bojanowski, P., Puhersch, C., & Joulin, A. (2017). Advances in pre-training distributed word representations. *arXiv preprint arXiv:1712.09405*.
- [12] Rahutomo, F., Kitasuka, T., & Aritsugi, M. (2012, October). Semantic cosine similarity. In *The 7th International Student Conference on Advanced Science and Technology ICAST* (Vol. 4, No. 1)