

Jutisi: Jurnal Ilmiah Teknik Informatika dan Sistem Informasi
 Jl. Ahmad Yani, K.M. 33,5 - Kampus STMIK Banjarbaru
 Loktabat – Banjarbaru (Tlp. 0511 4782881), e-mail: puslit.stmikbjb@gmail.com
 e-ISSN: 2685-0893
 p-ISSN: 2089-3787

Sistem Rekomendasi Pekerjaan Menggunakan *Content Based Similarity*

Abdur Ro'uf^{1*}, Yuliana Melita Pranoto², dan Endang Setyati³

Departemen Informatika, Institut Sains Dan Teknologi Terpadu Surabaya, Surabaya, Indonesia
 *e-mail *Corresponding Author*: abdurrouf.ar34@gmail.com

Abstract

Based on data from the Central Statistics Agency (BPS) it is stated that the data for people who did not have a job from August 2019 to August 2021 recorded an increase from 7104.42 to 9102.05, meaning that within 2 years people who did not have a job had increased significantly. This is caused by one of the factors, namely finding information on job vacancies which is difficult, users still have to choose one job at a time in accordance with their field of knowledge. By building a job recommendation system, users will find it easier to find suitable job information, the data used is obtained from 1120 (one thousand one hundred and twenty) alumni data which includes academic grades, non-academic scores, positions and companies obtained from alumni data from the Institute of Technology and Business Widya Gama Lumajang. Using a content-based similarity algorithm with machine learning techniques using the MLP classifier feature and several trials using different parameters in each experiment. In each experiment the researcher used 10 (ten) samples. The results of this trial the machine learning feature of the MLP classifier can be concluded to be able to provide an accuracy of 81% with a precision value of 0.77, a recall of 0.81 and an f1-score of 0.76. The results of this study are used by users or fresh graduates to get job recommendations in accordance with their field of study.

Keywords: *Content Based Similarity; Interaciton Based Relation; Job Recommendation System*

Abstrak

Berdasarkan data Badan Pusat Statistik (BPS) menyebutkan bahwa data orang yang tidak memiliki pekerjaan dari agustus 2019 sampai dengan agustus 2021 tercatat naik dari angka 7.104,42 menjadi 9.102,05 artinya dalam kurun 2 tahun orang yang tidak memiliki pekerjaan mengalami kenaikan secara signifikan. Hal ini disebabkan oleh salah satu faktor yaitu mencari informasi lowongan pekerjaan yang sulit pengguna masih harus memilih satu per satu pekerjaan yang sesuai dengan bidang ilmunya. Dengan membangun sistem rekomendasi pekerjaan pengguna akan lebih mudah menemukan informasi pekerjaan yang sesuai, data yang digunakan diperoleh dari data alumni sebanyak 1.120 (seribu seratus dua puluh) yang mencakupi nilai akademik, nilai non akademik, jabatan dan perusahaan yang diperoleh dari data alumni Institut Teknologi Dan Bisnis Widya Gama Lumajang. Menggunakan algoritma *content-based similarity* dengan teknik *machine learning* fitur *MLP classifier* dan beberapa kali uji coba menggunakan parameter yang berbeda-beda pada setiap percobaannya. Pada setiap percobaan peneliti memakai 10 (sepuluh) *sample*. Hasil dari uji coba ini machine learning fitur *MLP classifier* dapat disimpulkan mampu memberikan akurasi sebesar 81% dengan nilai *precision* 0.77 *recall* 0.81 dan *f1-score* 0.76. Hasil penelitian ini digunakan oleh pengguna atau *fresh graduate* untuk mendapatkan rekomendasi pekerjaan yang sesuai dengan bidang ilmu.

Kata Kunci: *Content Based Similarity; Interaciton Based Relation; Sistem Rekomendasi Pekerjaan*

1. Pendahuluan

Berdasarkan data Badan Pusat Statistik (BPS) menyebutkan bahwa data orang yang tidak mempunyai pekerjaan dari agustus 2019 sampai dengan agustus tahun 2021 tercatat naik dari angka 7.104 menjadi 9.102 artinya dalam kurun waktu 2 tahun orang yang tidak mempunyai pekerjaan bertambah 1.997 orang [1] data tersebut berdasarkan pada <https://www.bps.go.id/> .

Pada tahun Pesatnya perkembangan informasi dalam masyarakat, sebagian besar informasi tersebut bersumber pada situs web [2]. Sebagian besar situs web lowongan pekerjaan menampilkan informasi lamaran pekerjaan yang sangat bervariasi. Mahasiswa harus memilih informasi lowongan pekerjaan yang sesuai dengan kebutuhan mereka seperti pekerjaan yang sesuai dengan bidang ilmunya. Proses seperti inilah yang membuat pencari lowongan pekerjaan menjadi proses yang sangat panjang dan berdampak pada minat mencari informasi lowongan pekerjaan.

Berdasarkan data yang diambil dari halaman berita Databoks Kata Data [3] menurut Badan Pusat Statistik mahasiswa sarjana strata satu (S 1) yang menggugur pada tahun 2021 yaitu laporan yang diambil pada bulan februari mencapai hampir 1 juta orang yaitu lebih tepatnya mencapai 999.543 (Sembilan ratus Sembilan puluh Sembilan ribu lima ratus empat puluh tiga) orang. Hal ini diakibatkan oleh pandemic covid-19 yang menyebabkan krisis ekonomi sehingga sejumlah perusahaan mengurangi pekerjanya demi menghindari kerugian pada keuangan perusahaan tersebut. Dampak pada lulusan sarjana ialah berkurangnya lapangan pekerjaan pada masa pandemi yang semakin sulit dicari, dan terlebih lagi kebijakan pemerintah mengenai larangan tatap muka, mengurangi mobilitas diluar ruangan dan mengharuskan masyarakat Indonesia untuk berdiam dirumah sampai batas waktu yang tidak ditentukan hal ini berlaku di seluruh wilayah Indonesia, mahasiswa harus menggunakan perangkat komputer atau handphone berbasis android untuk mencari informasi lowongan pekerjaan yang sesuai dengan kriterianya. Dengan begitu mahasiswa masih harus memilih informasi lebih lagi untuk mendapatkan kriteria pekerjaan yang sesuai dengan bidangnya.

Sistem rekomendasi dapat menjadi salah satu pilihan untuk dimanfaatkan [4] sehingga pengguna tidak mengalami kelebihan informasi atau disebut overload information karena sistem ini dibangun untuk mampu merekomendasikan content atau item yang relevan dan sesuai dengan yang diinginkan pengguna, seperti dalam penelitian yang dilakukan oleh Muhtar Yusuf [5].

2. Tinjauan Pustaka

Pada tahun 2013 [6] penelitian yang dilakukan bertujuan untuk mencari lowongan pekerjaan sesuai seiring bertambahnya informasi pada situs web, maka orang-orang harus lebih teliti memilih pekerjaan. Metode yang digunakan dalam masalah ini ialah *hybrid* dan *graph modelling* bertujuan untuk mendapatkan hasil rekomendasi pekerjaan yang sesuai dan tepat untuk semua orang, dalam pemodelan peneliti menggunakan *interaction based relations* dan *content base* [6]. Pertama *Content based similarity* membangun hubungan dua arah berdasarkan hipotesis yang jika ada kecocokan pada pencari kerja dan pengguna yang mempunyai riwayat pernah menyukai hal yang sama atau biasa disebut dengan item. Kedua, dua entitas dari grup yang sama dihubungkan dengan relasi dua arah jika memiliki data diri atau profil yang serupa maka cenderung mempunyai minat yang sama. Sistem ini berbasis konten dengan membandingkan profil dua entitas. Saat menghitung kedua profil, sistem ini akan membandingkan dua entitas grup yang sama dengan bidang yang sesuai. Peneliti memasukkan atribut untuk menghitung sebagai poin penting untuk menghitung kemiripan data yang ada. Berikut output yang dihasilkan dalam penelitian ini *hybrid* 0,8 *profil similarity* 0,8 dan *collaborative filtering* 0,6. Nilai akurasi yang dihasilkan pada penelitian ini yaitu sebesar 80% untuk memberikan rekomendasi pekerjaan

Pada tahun 2015 [7] melakukan penelitian yang didasari oleh latar belakang tentang jumlah besar informasi dari berbagai sumber seperti koran, situs web, lagu, film, buku atau bahkan informasi lowongan pekerjaan. *Job recommendation system* banyak dikaji dari berbagai aspek seperti merangkum kategori-kategori yang mencakup manfaat dan keuntungan menggunakan *job recommendation system* [6]. Penelitian ini melakukan pendekatan dengan algoritma *hybrid recommendation* dan FoDRA [7] salah satunya memakai *content-based*, hasil menunjukkan dengan total sample yang diambil sebanyak 5 nilai teratas menunjukkan kesesuaian nilai paling rendah yaitu 4,171573 dan kesesuaian nilai paling tinggi mencapai 10,585786.

Pada tahun 2017 [8] dalam penelitiannya, industri sangat bergantung pada sistem penyaringan informasi yang dapat membantunya memberikan hasil rekomendasi yang tepat. Menurut [8] karena peningkatan data ini, daftar pekerjaan dan situs layanan rekrutmen mengalami kualitas dan kinerja. Dalam keadaan seperti ini baik pekerja maupun industri yang memberikan lowongan pekerjaan harus menyeleksi sebelum memilih dan menetapkan

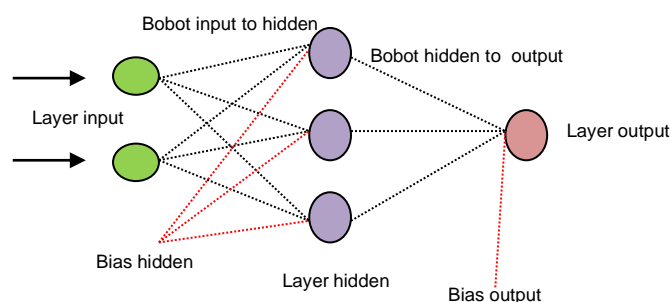
dikarenakan data informasi begitu banyak dan perlu menghasilkan informasi daftar pekerjaan yang sesuai. Tujuan dari penelitian ini ialah peneliti menyajikan prototipe implementasi baru dari algoritma rekomendasi yang menerapkan teknik pemodelan topik *Laten Dirischlet Allocation* (LDA) dan diantaranya peneliti menggunakan algoritma *content based*. Metrik pengukuran tingkat akurasi digunakan untuk mengukur kinerja sistem seperti presisi, *recall* dan *F1-measure* [9]. Pendekatan *content-based recommendation* menggunakan pemodelan topik LDA mampu mencapai pada rata-rata 95%.

Tahun 2021 [10] dalam penelitiannya, sebuah sistem rekomendasi pekerjaan ada beberapa atribut yang digunakan seperti jenis pekerjaan, tingkat pendidikan dan skill, sistem rekomendasi ini dirancang dengan memanfaatkan peringkat dan mencocokkan skill yang dibutuhkan oleh perusahaan. [11] Menggunakan fitur dan rating untuk membedakan antara pengguna atau informasi yang dibutuhkan pada penelitiannya langkah-langkah seperti ini dapat disebut sebagai klasifikasi. Teknik sistem rekomendasi pekerjaan terdiri dari berbagai jenis berdasarkan input data yang dibutuhkan [10] diantaranya *recommender system*, *content based filtering* dengan hasil yang di dapatkan sangat baik, tingkat akurasi digambarkan dengan menggunakan level 1 – 5 dengan hasil yang didapatkan mencapai level 4 untuk tingkat akurasinya.

Dengan beberapa tinjauan pustaka di atas, sistem rekomendasi banyak digunakan dalam membantu seseorang mencari informasi pekerjaan. Objek-objek pendukung dalam melakukan uji coba dilakukan oleh peneliti terdahulu dapat dikatakan kurang, penulis bermaksud menggunakan beberapa objek pendukung yang lebih banyak dengan maksud agar tingkat akurasi terhadap rekomendasi dari sistem tersebut lebih maksimal sehingga dapat menghasilkan sistem baru yang dapat digunakan dengan maksimal. Algoritma *content based similarity* dengan *fitur library sklearn MLP Classifier* dapat bekerja secara maksimal dengan kompleksitas dataset lebih banyak dikarenakan *MLP Classifier* sudah terhubung dengan *Neural Network*.

3. Metodologi

Pada penelitian ini menggunakan *content-based similarity* yaitu dapat dikatakan model pengembangan dari sistem-sistem sebelumnya, *content-based similarity* memusatkan kemampuan dalam *find of the similar data* dengan machine learning menggunakan fitur *library sklearn* yaitu *MLP Classifier*. *MLP Classifier* yang sudah terhubung ke dalam *Neural Network* ialah mempunyai tugas dasar untuk klasifikasi tidak seperti algoritma klasifikasi lain seperti *Support Vectors Machine* atau *Baive Bayes Classifier* [12]. *MLP Classifier* lebih mudah di implementasikan dari pada algoritma klasifikasi *scikit-learn* lainnya [12]. Cara kerja *MLP Classifier*, input data pada sistem kemudian sistem akan menentukan bobot sebelum di test dengan tiga lapis *hidden_layer_size* atau dapat dikatan dengan bias *hidden* proses ini untuk menguji sistem kerja algoritma tersebut pada lapisan keberapa algoritma tersebut menemukan akurasi yang maksimal selanjutnya setelah sistem menemukan data yang paling tinggi nilainya akan menjadi *output* atau hasil. Berikut gambar ilustrasi *MLP Classifier*.



Gambar 1. Ilustrasi *MLP Classifier*

1) *Content Based Similarity*

Content based similarity ialah algoritma yang akan digunakan pada penelitian ini untuk menentukan klasifikasi dan menentukan prediksi regresi. *Content based similarity* diharapkan

dapat memberikan hasil yang baik dalam memberikan rekomendasi pekerjaan dalam penelitian ini.

a). *Content Based Similarity*

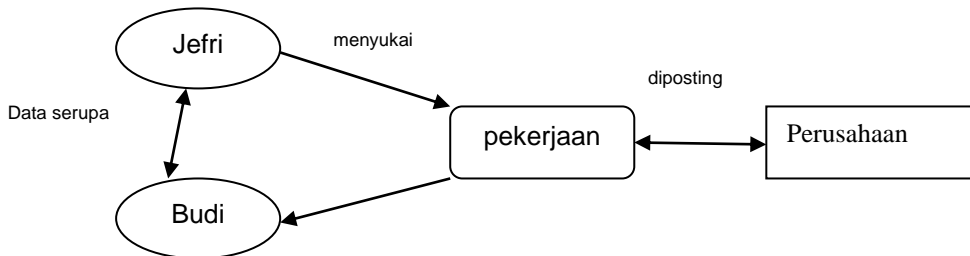
Content based similarity adalah sebuah sistem yang membangun hubungan dua arah berdasarkan kemiripan data [13], metode *content based similarity* memberikan rekomendasi dari beberapa objek memiliki atribut atau data yang mirip, objek direkomendasikan dan objek yang telah dipilih. metode ini tidak memiliki faktor bergantung pada situasi seperti item tersebut merupakan item baru yang belum pernah dipilih ataupun bukan item baru. Metode ini mempunyai sifat *user independence*. Cara kerja sistem *content-based similarity* yaitu seperti seorang yang memesan makanan pada suatu aplikasi maka sistem tersebut akan memberikan tawaran lain yang menurut sistem tersebut memiliki data kemiripan yang sama. Rumus yang biasa digunakan untuk menghitung *similarity* ialah algoritma *pearson correlation* [14]. Berikut bentuk persamaannya:

$$sim(i, j) = \frac{\sum_{u=1}^m R_{u,i} - \bar{R}_i \quad R_{u,j} - \bar{R}_j}{\sqrt{\sum_{u=1}^m R_{u,i} - \bar{R}_i \quad 2 \quad \sum_{u=1}^m R_{u,j} - \bar{R}_j \quad 2}} \dots\dots\dots (1)$$

- Sim (i,j)* = nilai similarity antara item *i* dan *j*
- m* = jumlah total user yang memberikan rating item *k* dan item *i*
- R_i* dan *R_j* = rata-rata pada item *i* dan item *j*
- R_{ui}* dan *R_{uj}* = rating yang diberikan oleh user *u* kepada item *i* dan item *j*

b). *Interaction Based Relation*

Interaction based relation adalah algoritma yang membangun interaksi dua arah dapat diartikan interaksi dua arah seperti antara pengguna pekerja dan pencari lowongan pekerjaan [15], saat pengguna menyatakan minat terhadap suatu item tertentu maka sistem tersebut akan merekam datanya dan sistem akan memberikan suatu item yang sama kepada pengguna yang lain dikarenakan pengguna sebelumnya mempunyai kemiripan data dengan pengguna saat ini. Contoh ilustrasi dapat dilihat pada gambar 1 berikut ini.



Gambar 2. Ilustrasi *Interaction Based Relation*

c). *Sistem Rekomendasi*

Sistem rekomendasi ialah sebuah sistem penyaringan informasi yang memberikan atau menampilkan data informasi pribadi kepada pengguna berdasarkan minat, relevansi informasi dan lain-lain[10]. Sistem ini sudah banyak digunakan untuk mendapatkan sebuah informasi baik untuk merekomendasikan artikel, film, kuliner dan termasuk pekerjaan [16]. Dalam suatu percobaan satu data pelanggan dan satu data pekerja, jika pengguna akan mendapatkan rekomendasi pekerjaan itu berdasarkan peringkat tapi jika pekerja tidak mempunyai peringkat maka itu tidak akan direkomendasikan kepada pengguna. Jadi, untuk mengatasinya peneliti menggunakan *content-based similarity*. Pendekatan menggunakan *content-based similarity* dengan kumpulan data set serta beberapa komponen lainnya yang diangkat dapat menjadi parameter perhitungan kemiripan data set dengan pengguna saat ini. Misalnya sistem rekomendasi akan merekomendasikan suatu pekerjaan yang memiliki kemiripan data pribadi.

Berikut adalah gambaran proses *content-based similarity*:

1. Identifikasi permasalahan

Identifikasi permasalahan ialah langkah pertama yang harus diputuskan dalam menentukan permasalahan dengan begitu peneliti dapat mengangkat topik yang akan dibahas.

2. Pengumpulan data set
Pengumpulan data set dilakukan sebagai modal yang akan digunakan untuk melatih algoritma dalam memberikan hasil. Data yang dikumpulkan berupa data alumni yang sudah bekerja seperti data pekerjaan, jabatan, perusahaan, skill dan nilai akademik. Pengumpulan data dilakukan dengan cara melacak data alumni yang sudah bekerja di *tracer study* dan juga dapat mengirimkan google form untuk mendapatkan data.
3. Data set dirubah ke dalam bentuk csv.
4. Menggunakan algoritma di dalam sistem rekomendasi seperti *content based similarity* dan *interaction based relation*.
5. Membuat model sistem pembelajaran untuk melakukan klasifikasi seperti *Model Learning Perceptron (MLP) Classifier*.
6. Mengukur nilai kemiripan dengan menggunakan *precision, recall* dan *F1 score*.
7. Menentukan lima angka teratas atau tertinggi untuk ditentukan sebagai hasil dari sistem rekomendasi.

2) Dataset

Pada penelitian ini, data yang digunakan ialah data alumni Institut Teknologi Dan Bisnis Widya Gama Lumajang sebanyak 2.000 (dua ribu) data alumni yang telah bekerja. Data yang dapat dipakai sebanyak 1.120 (seribu seratus dua puluh) dan sebanyak 880 (delapan ratus delapan puluh) data yang tidak dapat dipakai dikarenakan tidak tersedianya kelengkapan nilai akademik dan nilai non akademik. Data yang digunakan diantaranya nilai akademik, nilai non akademik, perusahaan, dan jabatan. Adapun data-data yang dibutuhkan sebagai berikut :

1. Nomor ID yaitu identitas pada *dataset*.
2. Jabatan ialah kedudukan pada suatu perusahaan.
3. Perusahaan ialah tempat yang akan digunakan sebagai rekomendasi kepada data tes.
4. Skill ialah nilai dari non akademik yang berhubungan dengan keahlian lain selain bidang akademik.
5. Nilai akademik ialah data nilai-nilai mata kuliah.

Tabel 1. Dataset Alumni, Perusahaan dan Pekerjaan

Nomor ID	Jabatan	Level Jabatan	Perusahaan	Level Perusahaan	Tingkat Kecocokan	Skill Perpajakan	Skill Komputer	Skill Marketing	Skill Laporan Keuangan	Skill Speaking	Bank Dan Lembaga Keuangan Lainnya	Akuntansi Biaya	Komunikasi Bisnis	Manajemen SDM I	Sistem Informasi Manajemen	Manajemen Pemasaran I	Manajemen Keuangan I
216122349	TELLER	1	PT BPR DHARMA INDIRA CBG YOSOWILANGUN	3	4	2	3	2	2	4	87	87	90	87	88	89	89
216122350	CUSTOMER SERVICE	2	BRI	5	4	3	4	4	4	4	91	87	90	92	91	87	80
216122351	ADMIN KREDIT	2	PT BPR SAA CBG KENCONG SUKODOLO	4	3	2	3	2	3	2	76	69	85	88	92	88	76
216122352	KASIR	1	TOKO SARI SOFA	1	4	1	1	1	1	2	80	85	86	86	85	84	88
216122353	TELLER	1	PT BPR DHARMA INDIRA	3	5	4	3	4	4	5	90	88	85	84	94	84	93
216122354	ADMIN KEUANGAN	3	ADIRA FINANCE	3	5	3	3	2	5	2	93	90	89	92	89	95	89
216122355	TELLER	1	PT BPR SAA CBG KENCONG	4	4	2	3	4	3	4	90	88	90	89	90	88	95
216122356	SALES PROMOTION	2	CV ALAM INDO	3	4	1	1	3	1	3	88	83	84	85	86	87	88
216122357	TELLER	1	PT BNI KC LUMAJANG	5	5	3	4	4	3	4	88	85	89	90	85	95	89
216122358	KASIR	1	JAMU GEJLUK	1	2	1	1	2	1	2	82	79	79	81	79	77	80
216122359	ADMIN PEMBAYARAN	2	BANK BTPN KCP LUMAJANG	3	4	3	4	4	4	4	85	88	92	85	75	87	88
216122360	ADMIN PERPAJAKAN	4	PT MUSTIKA TAMA GROUP	4	5	4	4	2	4	3	85	89	91	88	78	89	90
216122361	ADMIN KEUANGAN	3	UPT FARMASI LUMAJANG	2	4	2	3	3	4	3	79	79	85	89	92	85	86
216122362	ADMIN KEUANGAN	3	PT MATAHARI JAYA INTERNASIONAL	5	4	2	4	3	4	3	93	85	89	91	84	85	89
216122364	ADMIN KEUANGAN	3	CV UNGGUL MAKMUR SELAJHTERA	2	3	2	3	3	4	3	91	78	85	89	91	85	77
216122365	TELLER	1	PT BPR ADIRA FINANCE	5	3	2	3	4	3	4	93	85	89	91	87	90	89
216122366	ADMIN KEUANGAN	3	PT INDOLAKTO	2	3	2	3	3	4	3	85	84	90	85	78	89	92
216122367	KARYAWAN PRODUKSI	1	PT MUSTIKA TAMA GROUP	4	4	1	1	1	1	1	77	72	73	71	70	69	75
216122368	HRD	4	CV LUMAJANG MAKMUR BERSAMA	2	2	2	2	3	3	3	85	83	92	85	76	89	88

b. Preprocessing Data

Data yang akan digunakan pada proses penghitungan akan dilakukan *preprocessing* terlebih dahulu, seperti pelengkapan data yang kosong akibat penarikan data dari sistem akademik. Data yang kosong biasanya muncul dengan kata *null*, data tersebut dapat kita lengkapi kembali dengan mengecek data tersebut di sistem akademik Institut Teknologi Dan Bisnis Widya Gama Lumajang.

Kemudian menambahkan level perusahaan, level jabatan dan tingkat kecocokan dengan memberikan label 1 (satu) sampai dengan 5 (lima) sesuai dengan tingkat atau status perusahaan di masyarakat, tingkat jabatan pada perusahaan, dan tingkat kecocokan alumni

yang bekerja saat itu. Begitu juga pada nilai non akademik yang dinamakan dengan skill, level pada skill yaitu di ambil pada nilai kegiatan non akademik yang diperoleh pada saat masih belajar dengan mengekstraksi nilai tersebut dengan level 1 (satu) sampai dengan 5 (lima).

Tabel 2. Level Tingkat Kecocokan

Level Tingkat Kecocokan	Angka
SANGAT COCOK	5
COCOK	4
NETRAL	3
KURANG COCOK	2
TIDAK COCOK	1

Tabel 3. Level Tingkat Perusahaan

Level Tingkat Perusahaan	Angka
PERUSAHAAN SANGAT BESAR	5
PERUSAHAAN BESAR	4
PERUSAHAAN BIASA	3
PERUSAHAAN KECIL	2
PERUSAHAAN SANGAT KECIL	1

Tabel 3. Level Skill

Level Skill	Angka
BAIK SEKALI	5
BAIK	4
BIASA SAJA	3
BURUK	2
BURUK SEKALI	1

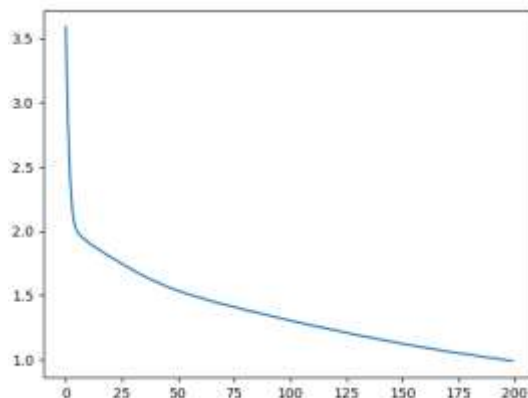
Tabel 4. Level Jabatan

Level Jabatan	Angka
BAIK SEKALI	5
BAIK	4
BIASA SAJA	3
BURUK	2
BURUK SEKALI	1

c. *Processing Data*

Dataset yang sudah siap untuk di uji coba kemudian memasukkannya kedalam model penelitiannya dengan komposisi 35% sebagai *test* 65% sebagai *training*. Berikut tahapan-tahapan yang dilakukan.

1. Membuat model yang akan digunakan untuk melakukan klasifikasi dengan *dataset training*.
2. Data yang digunakan sebagai sample sebanyak 10 (sepuluh) data.
3. Karena memakai MLP Classifier maka peneliti menggunakan *max_iter = 200* *hidden_layer_size = 100*. Hasil pelatihan model MLP Classifier atau tingkat *loss iterasi* dapat dilihat pada gambar dibawah ini :



Gambar 3. *Loss Iterasi*

4. Selanjutnya gunakan *numpy* untuk menghitung argumen maksimum dari *array y_pred* dengan sumbu atau *axis 1*.
5. Kemudian dilanjutkan dengan menghitung argumen maksimum dari *array*, dan juga membandingkan nilai prediksi dengan nilai aktual.
6. selanjutnya untuk mengevaluasi kerja, dapat menggunakan *library scikit-learn* untuk mengevaluasi kinerja model klasifikasi berupa matrik evaluasi klasifikasi termasuk *accuracy, precision, recall dan f1-score*.

6. Hasil Dan Pembahasan

Hasil yang didapatkan pada saat uji coba menggunakan teknik *machine learning* mencapai tingkat akurasi yang baik, hasil tersebut ditampilkan pada gambar 2. Beberapa percobaan dilakukan bertujuan mendapatkan hasil dengan akurasi yang baik, percobaan pertama dilakukan menggunakan parameter dengan *test_size* sebanyak 25% atau sebanyak 280 (dua ratus delapan puluh) data dengan *hidden_layer_size* 200, 300, 100 dan iterasi maksimal 200 hasil yang didapatkan menunjukkan *weighted avg* masing-masing pada *precision* 1.00, *recall* 1.00 dan *f1-score* 1.00 hasil tersebut sangat baik namun terdapat kinerja kurang maksimal pada saat *training*, hasil *training* menunjukkan iterasi tidak sampai pada data ke 280. Percobaan kedua, dilakukan menggunakan parameter dengan *test_size* sebanyak 35% atau sebanyak 392 (tiga ratus sembilan puluh dua) data dengan *hidden_layer_size* 100 dan iterasi maksimal 200 menunjukkan hasil *weighted avg* masing-masing pada *precision* 0.77, *recall* 0.81 dan *f1-score* 0.76. Hasil percobaan pertama dan percobaan kedua dapat dilihat pada Tabel 6 dan Tabel 7 berikut:

	Precision	Recall	F1-score	support
0	1.00	1.00	1.00	34
1	1.00	1.00	1.00	5
2	1.00	1.00	1.00	45
3	1.00	1.00	1.00	156
4	1.00	1.00	1.00	40
Accuracy			1.00	280
Macro avg	1.00	1.00	1.00	280
Weighted avg	1.00	1.00	1.00	280

Tabel 6. Hasil Percobaan

	Precision	Recall	F1-score	support
0	1.00	1.00	1.00	39
1	0.00	0.00	0.00	6
2	0.50	0.10	0.16	63
3	0.77	0.99	0.86	221
4	0.96	0.84	0.90	63
Accuracy			0.81	392
Macro avg	0.65	0.59	0.58	392
Weighted avg	0.77	0.81	0.76	392

Tabel 7. Hasil Percobaan

Dengan hasil yang ada pada tabel di atas maka peneliti menggunakan proses pada percobaan kedua, yang dinilai lebih baik digunakan untuk menghitung kemiripan data. Dapat dilihat model MLP Classifier memberikan hasil yang baik pada dataset yang seimbang dengan hasil 81% yang artinya model ini dapat mengklasifikasikan dengan baik dari seluruh data, *precision* menunjukkan hasil 77% ialah model ini dapat menghasilkan rasio antara jumlah prediksi benar dan juga dengan jumlah prediksi positif, hasil *recall* sebesar 81% artinya model MLP Classifier tersebut mampu memberikan informasi tentang kasus positif dengan baik sebesar 81%, *f1-score* sebesar 76% juga dapat dikatakan model ini dapat memberikan keseimbangan antara *precision* dan *recall* artinya model ini dapat bekerja secara optimal. *Content based similarity* dengan model MLP Classifier dapat digunakan untuk mendapatkan suatu rekomendasi.

5. Simpulan

Berdasarkan dari hasil penelitian yang telah dilakukan bahwa penerapan data mining klasifikasi menggunakan *content-based similarity* untuk mendapatkan rekomendasi pekerjaan, peneliti menggunakan dua kali tahap percobaan dengan menerapkan parameter yang berbeda. Percobaan ini menggunakan teknik *machine learning* dengan fitur MLP Classifier, data yang digunakan sebanyak 1.120 (seribu seratus dua puluh) terdiri dari komponen jabatan, perusahaan, nilai non akademik dan nilai akademik. Tidak hanya itu peneliti menambahkan komponen level jabatan, level perusahaan dan tingkat kecocokan.

Percobaan pertama, peneliti menggunakan parameter dengan komposisi *test_size* 25% atau sebanyak 280 (dua ratus delapan puluh) data, *hidden_layer_size* 200, 300, 100 dan iterasi maksimal 200 dengan komposisi tersebut percobaan pertama mencapai tingkat akurasi sebesar 100% dengan nilai *precision* 1.00 *recall* 1.00 dan *f1-score* 1.00 atau sangat baik namun pada *training*, proses iterasi tidak dapat menyelesaikan sampai 200 maka peneliti memutuskan untuk membuat satu percobaan lagi.

Percobaan kedua, peneliti menggunakan parameter dengan komposisi *test_size* 35% atau sebanyak 392 (tiga ratus sembilan puluh dua) data, *hidden_layer_size* 100 dan iterasi maksimal 200 dengan komposisi seperti itu hasil dari percobaan tersebut mampu mencapai tingkat akurasi yaitu sebesar 81% dengan nilai *precision* 0.77 *recall* 0.81 dan *f1-score* 0.76.

Kesimpulan pada penelitian ini dengan menggunakan algoritma *content-based similarity*, peneliti harus melakukan percobaan lebih dari satu kali untuk mendapatkan hasil yang maksimal dengan menggunakan teknik *machine learning* dengan fitur MLP *Classifier* dan mencoba dengan parameter yang berbeda akan memberikan hasil yang maksimal.

Pada penelitian selanjutnya diharapkan peneliti dapat memanfaatkan metode lainnya seperti metode *hybrid* dari beberapa metode untuk membandingkan tingkat akurasi yang dihasilkan sehingga peneliti dapat mengetahui metode yang baik untuk digunakan.

Daftar Referensi

- [1] Badan Pusat Statistik, "Data Badan Pusat Statistik (BPS) dari bulan agustus 2019 sampai dengan bulan agustus tahun 2021 tercatat naik dari angka 7 104,42 menjadi 9102,05 artinya dalam kurun waktu 2 tahun orang yang tidak mempunyai pekerjaan bertambah 1997,63." <https://www.bps.go.id/>.
- [2] X. Guo, H. Jerbi, dan M. P. O'Mahony, "An Analysis Framework for Content-based Job Recommendation".
- [3] Databoks Katadata, "BPS: Sarjana yang Menganggur Hampir 1 Juta Orang pada Februari 2021." [Daring]. Tersedia pada: <https://databoks.katadata.co.id/>
- [4] F. Nurfalah, Asriyanik, dan A. Pambudi, "Sistem Rekomendasi Event Online Menggunakan Metode Content Based Filtering," *Elkom J. Elektron. Dan Komput.*, vol. 15, no. 2, hlm. 271–279, Des 2022, doi: 10.51903/elkom.v15i2.736.
- [5] M. Yusuf dan A. Cherid, "Implementasi Algoritma Cosine Similarity Dan Metode TF-IDF Berbasis PHP Untuk Menghasilkan Rekomendasi Seminar," vol. 9, no. 1, 2020, [Daring]. Tersedia pada: <https://publikasi.mercubuana.ac.id/index.php/fasilkom/article/view/8830>
- [6] W. Hong, S. Zheng, H. Wang, dan J. Shi, "A Job Recommender System Based on User Clustering," *J. Comput.*, vol. 8, no. 8, hlm. 1960–1967, Agu 2013, doi: 10.4304/jcp.8.8.1960-1967.
- [7] N. D. Almalis, G. A. Tsihrintzis, N. Karagiannis, dan A. D. Strati, "FoDRA — A new content-based job recommendation algorithm for job seeking and recruiting," dalam *2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA)*, Corfu: IEEE, Jul 2015, hlm. 1–7. doi: 10.1109/IISA.2015.7388018.
- [8] S. Bansal, A. Srivastava, dan A. Arora, "Topic Modeling Driven Content Based Jobs Recommendation Engine for Recruitment Industry," *Procedia Comput. Sci.*, vol. 122, hlm. 865–872, 2017, doi: 10.1016/j.procs.2017.11.448.
- [9] D. Behl, S. Handa, dan A. Arora, "A bug Mining tool to identify and analyze security bugs using Naive Bayes and TF-IDF," dalam *2014 International Conference on Reliability Optimization and Information Technology (ICROIT)*, Faridabad, Haryana, India: IEEE, Feb 2014, hlm. 294–299. doi: 10.1109/ICROIT.2014.6798341.
- [10] M. Purkar, O. Joshi, A. Salape, A. Patil, V. Kulkarni, dan P. Futane, "Recommendation System for Workers & Customers for Informal Jobs," dalam *SSRN Electronic Journal*, Vishwakarma Institut of Information Technology: Vishwakarma Institut of Information Technology, 2021, hlm. 1–6. doi: 10.2139/ssrn.3833762.
- [11] J. Wintrode, G. Sell, A. Jansen, M. Fox, D. Garcia-Romero, dan A. McCree, "Content-based recommender systems for spoken documents," dalam *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, South Brisbane, Queensland, Australia: IEEE, Apr 2015, hlm. 5201–5205. doi: 10.1109/ICASSP.2015.7178963.
- [12] Nurtiyasari, Devi dan Syarif, Mulkan, "Belajar Algoritma Multi Layer Perceptron Continue reading at Belajar Algoritma Multi Layer Perceptron | Softscients," *Belajar Algoritma Multi Layer Perceptron Continue reading at Belajar Algoritma Multi Layer Perceptron | Softscients*, 2020. <https://softscients.com/2020/11/05/belajar-algoritma-multi-layer-perceptron/>
- [13] R. H. Mondri dan A. Wijayanto, "Recommendation System With Content-Based Filtering Method For Culinary Tourism In Mangan Application," *ITSMART: Jurnal Teknologi dan Informasi*, vol. 8, no. 2, hlm. 65–72, 2019, doi: <https://doi.org/10.20961/itsmart.v8i2.35008>.
- [14] M. Irfan, Cahyani, Andriani D, dan Hastarita, Fika, "Sistem Rekomendasi: Buku Online Dengan Metode Collaborative Filtering," *Jurnal Teknologi Technoscintia*, vol. 7, no. 1, hlm. 76–84, 2014, doi: <https://doi.org/10.34151/technoscintia.v7i1.612>.
- [15] S. Dong, Z. Lei, P. Zhou, K. Bian, dan G. Liu, "Job and Candidate Recommendation with Big Data Support: A Contextual Online Learning Approach," dalam *GLOBECOM 2017 -*

- 2017 *IEEE Global Communications Conference*, Singapore: IEEE, Des 2017, hlm. 1–7. doi: 10.1109/GLOCOM.2017.8255006.
- [16] I. W. Jepriana dan S. Hanief, “Analisis Dan Implementasi Metode Item-Based Collaborative Filtering Untuk Sistem Rekomendasi Konsentrasi Di Stmik Stikom Bali,” *Jurnal Nasional Pendidikan Teknik Informatika: JANAPATI*, vol. 9, no. 2, hlm. 171–180, 2020, doi: <https://doi.org/10.23887/janapati.v9i2.23218>.