

**IMPLEMENTASI *IMPROVED SQRT-COSINE SIMILARITY*  
UNTUK PEMERINGKATAN RESUME BERDASARKAN  
KUALIFIKASI LOWONGAN KERJA**

**SKRIPSI**

Untuk memenuhi sebagian persyaratan  
memperoleh gelar Sarjana Komputer

Disusun oleh:  
Khansa Salsabila Sangdiva Laksono  
NIM: 215150201111068



TEKNIK INFORMATIKA  
DEPARTEMEN TEKNIK INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS BRAWIJAYA  
MALANG  
2025

## **PERSETUJUAN**

**IMPLEMENTASI IMPROVED SQRT-COSINE SIMILARITY UNTUK PEMERINGKATAN  
RESUME BERDASARKAN KUALIFIKASI LOWONGAN KERJA**

**SKRIPSI**

Diajukan untuk memenuhi sebagian persyaratan  
memperoleh gelar Sarjana Komputer

Disusun Oleh :  
Khansa Salsabila Sangdiva Laksono  
NIM: 215150201111068

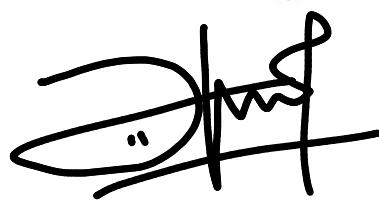
Skripsi ini telah disetujui pada  
24 Juni 2025  
Telah diperiksa dan disetujui oleh:

Dosen Pembimbing 1



Rizal Setya Perdana, S.Kom., M.Kom., Ph.D.  
NIK: 201603 910118 1 001

Dosen Pembimbing 2



Ir. Indriati, S.T., M.Kom.  
NIP: 19831013 201504 2 002

## **PERNYATAAN ORISINALITAS**

Saya menyatakan dengan sebenar-benarnya bahwa sepanjang pengetahuan saya, di dalam naskah skripsi ini tidak terdapat karya ilmiah yang pernah diajukan oleh orang lain untuk memperoleh gelar akademik di suatu perguruan tinggi, dan tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis disitasi dalam naskah ini dan disebutkan dalam daftar referensi.

Apabila ternyata didalam naskah skripsi ini dapat dibuktikan terdapat unsur-unsur plagiasi, saya bersedia skripsi ini digugurkan dan gelar akademik yang telah saya peroleh (sarjana) dibatalkan, serta diproses sesuai dengan peraturan perundang-undangan yang berlaku (UU No. 20 Tahun 2003, Pasal 25 ayat 2 dan Pasal 70).

Malang, 3 Juli 2025



A handwritten signature in black ink, appearing to read 'Khansa Salsabila Sangdiva Laksono', is placed next to the QR code.

Khansa Salsabila Sangdiva Laksono

NIM: 215150201111068

## PRAKATA

Puji syukur penulis panjatkan ke hadirat Allah SWT yang telah melimpahkan rahmat dan hidayah-Nya sehingga penulis dapat menyelesaikan skripsi yang berjudul “Implementasi *Improved Sqrt-Cosine Similarity* Untuk Pemeringkatan ResUME Berdasarkan Kualifikasi Lowongan Kerja”. Penulis menyadari bahwa dalam penyusunan skripsi tidak terwujud tanpa adanya dukungan, bimbingan, arahan, serta doa yang tiada hentinya dari berbagai pihak. Pada kesempatan kali ini penulis mengucapkan terima kasih sebesar-besarnya kepada:

1. Bapak Rizal Setya Perdana, S.Kom., M.Kom., Ph.D. selaku dosen pembimbing satu yang telah menyetujui dan mengarahkan penulis sehingga dapat menyelesaikan skripsi ini.
2. Ibu Ir. Indriati, S.T., M.Kom. selaku dosen pembimbing dua yang telah menyetujui dan membimbing dalam penulisan untuk pengeraaan skripsi ini.
3. Bapak Bayu Priyambadha, S.Kom., M.Kom., Ph.D. selaku Ketua Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Brawijaya.
4. Bapak Sabriansyah Rizqika Akbar, S.T., M.Eng., Ph.D. selaku Ketua Departemen Teknik Informatika Fakultas Ilmu Komputer Universitas Brawijaya.
5. Rajiv Maulana selaku validator dalam skripsi ini, serta seluruh rekan kerja penulis yang telah berkontribusi dalam memperluas wawasan dan pengetahuan penulis selama proses penelitian.
6. Ayah Tripinto Laksono, S.Kom. dan Bunda Dian Laksono selaku kedua orang tua penulis, Sangkaisar Laksono selaku adik penulis, dan seluruh keluarga penulis yang senantiasa memberikan dukungan, doa, dan motivasi sehingga dapat menyelesaikan skripsi ini.
7. P4OP Dinas Pendidikan Jakarta selaku penyelenggara beasiswa KJMU yang membantu penulis dalam menyelesaikan studi sarjana.
8. Seluruh teman tercinta penulis hingga saat ini yang telah menjadi teman diskusi selama proses penelitian, teman seperjuangan, serta sumber motivasi, terutama Salsabila Rachmayani, Kirana Alivia, Nathania Putri, Aidah Az Zahra, Raditya Atmaja, Roian Zain, Ade Arya, Nadhira Nurannisa, Saqina Salsabila, Ghania Tanziela, Gustav Ali, Emilia Putri, Farel Rakha, Aldiansyah, Dzaki Rafif, Bagas Antarino, Safia Putri, Rayshanda Yuwandina, Arkan, Alka, Faqih, Audrey, Aelissa, Dina, Kurnia, dan Zahra.

Malang, 25 Juni 2025

Penulis

khansalaksono@gmail.com

## **ABSTRAK**

**Khansa Salsabila Sangdiva Laksono, Implementasi *Improved Sqrt-Cosine Similarity* Untuk Pemeringkatan Resume Berdasarkan Kualifikasi Lowongan Kerja**

**Pembimbing: Rizal Setya Perdana, S.Kom., M.Kom., Ph.D. dan Indriati, Ir., S.T., M.Kom.**

Ketidaksesuaian antara kualifikasi pelamar dengan kebutuhan penyedia lowongan kerja dapat menjadi salah satu penyebab fenomena pengangguran. Penelitian ini menggunakan pendekatan representasi teks TF-IDF dan Word2Vec untuk implementasi perhitungan similaritas *Improved Sqrt-Cosine* (ISC) antara resume dengan kualifikasi lowongan kerja, memeringkat lima resume per kualifikasi lowongan kerja, dan dievaluasi hasilnya oleh seorang ahli dengan dua skenario yang melibatkan pemberian bobot pada setiap *section* dalam resume. Hasil penelitian ini menunjukkan keunggulan pada Word2Vec dengan ISC pada skenario tanpa bobot *section* dan Word2Vec dengan *Cosine Similarity* pada skenario dengan bobot *section*. TF-IDF dengan ISC menunjukkan performa terbaik dalam menghasilkan lima resume yang isiannya relevan dengan deskripsi lowongan kerja. Meskipun implementasi ISC dengan representasi teks Word2Vec unggul karena masih cukup mampu menangkap hubungan semantik kata kunci, tetapi kurang disarankan karena mengaburkan hubungan semantik asli akibat nilai absolut. Jika preferensi bobot *section* dapat menimbulkan bias karena kurang mencerminkan variasi preferensi rekruter pada umumnya, maka implementasi ISC dengan TF-IDF lebih disarankan untuk digunakan.

Kata kunci: similaritas resume, pemeringkatan, *improved sqrt-cosine*, spearman

## **ABSTRACT**

**Khansa Salsabila Sangdiva Laksono, The Implementation of Improved Sqrt-Cosine Similarity for Resume Ranking Based on Job Vacancy Qualifications**

**Supervisors: Rizal Setya Perdana, S.Kom., M.Kom., Ph.D. and Indriati, Ir., S.T., M.Kom.**

The mismatch between a job applicant's qualifications and the requirements of job providers can contribute to the phenomenon of unemployment. This research employs TF-IDF and Word2Vec text representation approaches to implement the Improved Sqrt-Cosine (ISC) similarity calculation between resumes and job vacancy qualifications, ranking the top five resumes per job qualification, and evaluating the results by an expert using two scenarios involving the weighting of resume sections. The results indicate that Word2Vec with ISC performs best in the scenario without section weighting, while Word2Vec with Cosine Similarity excels in the scenario with section weighting. TF-IDF with ISC demonstrates the best performance in generating the top five resumes with content relevant to the job description. Although the implementation of ISC with Word2Vec performs best because it is still quite capable of capturing the semantic relationships of keywords, it is less recommended due to the distortion of original semantic relationships caused by absolute value transformations. If section weighting preferences introduce bias by not reflecting the general preferences of recruiters, the implementation of ISC with TF-IDF is more recommended for use.

Keywords: resume similarity, ranking, improved sqrt-cosine, spearman

## DAFTAR ISI

PENGESAHAN .....	ii
PERNYATAAN ORISINALITAS .....	iii
PRAKATA.....	iv
ABSTRAK.....	v
ABSTRACT.....	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xiii
DAFTAR LAMPIRAN .....	xv
BAB 1 PENDAHULUAN .....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah .....	2
1.3 Tujuan.....	2
1.4 Manfaat .....	3
1.5 Batasan Masalah .....	3
1.6 Sistematika Pembahasan .....	3
BAB 2 LANDASAN KEPUSTAKAAN.....	5
2.1 Kajian Pustaka .....	5
2.2 Dasar Teori .....	7
2.2.1 Resume.....	7
2.2.2 Similaritas Teks.....	8
2.2.3 Pra-pemrosesan Teks .....	9
2.2.4 TF-IDF .....	10
2.2.5 Word2Vec.....	12
2.2.6 <i>Improved Sqrt-Cosine Similarity</i> .....	13
2.2.7 <i>Cosine Similarity</i> .....	14
2.2.8 <i>Human-Level Performance</i> .....	14
2.2.9 <i>Spearman Rank Correlation Coefficient (SRCC)</i> .....	15
BAB 3 METODOLOGI.....	16

3.1	Tipe Penelitian.....	16
3.2	Strategi Penelitian .....	16
3.3	Lokasi Penelitian.....	16
3.4	Metode Pengumpulan Data.....	16
3.5	Metode Analisis Data .....	17
3.6	Metode Evaluasi.....	17
3.7	Peralatan Pendukung .....	18
3.7.1	Perangkat Lunak ( <i>Software</i> ).....	18
3.7.2	Perangkat Keras ( <i>Hardware</i> ) .....	18
3.8	Perancangan Algoritma.....	18
BAB 4	PERANCANGAN .....	20
4.1	Deskripsi Umum .....	20
4.2	<i>Preprocessing</i> .....	20
4.2.1	Ekstraksi <i>Section</i> .....	20
4.2.2	<i>Preprocessing</i> Isian Resume .....	25
4.2.3	<i>Preprocessing</i> Penamaan <i>Section</i> .....	27
4.2.4	<i>Preprocessing</i> Kualifikasi Lowongan Kerja .....	29
4.3	Perhitungan Representasi Teks.....	30
4.3.1	TF-IDF .....	30
4.3.2	Word2Vec.....	33
4.4	Perhitungan Similaritas .....	36
4.4.1	<i>Improved Sqrt-Cosine Similarity</i> .....	36
4.4.2	<i>Cosine Similarity</i> .....	38
4.5	Perhitungan Korelasi .....	40
4.6	Perhitungan Relevansi dan Senioritas.....	42
4.7	Perhitungan Manual.....	44
4.7.1	Data Uji.....	44
4.7.2	Perhitungan Manual Ekstraksi <i>Section</i> .....	45
4.7.3	Perhitungan Manual <i>Preprocessing</i> Resume .....	46
4.7.4	Perhitungan Manual <i>Preprocessing</i> Kualifikasi Lowongan Kerja.....	59
4.7.5	Perhitungan Manual Representasi Teks .....	61

4.7.6	Perhitungan Manual Similaritas.....	94
4.7.7	Skenario Pengujian.....	101
BAB 5	IMPLEMENTASI.....	105
5.1	Implementasi Kode Program <i>Import Libraries</i> dan <i>Load Dataset</i> .....	105
5.2	Implementasi Kode Program <i>Preprocessing</i> Resume .....	108
5.3	Implementasi Kode Program <i>Preprocessing</i> Kualifikasi Lowongan Kerja .....	118
5.4	Implementasi Kode Program Representasi Teks TF-IDF .....	120
5.5	Implementasi Kode Program Representasi Teks Word2Vec	
	121	
5.6	Implementasi Kode Program Perhitungan Similaritas.....	123
5.6.1	Implementasi Kode Program <i>Improved Sqrt-Cosine Similarity</i> .....	123
5.6.2	Implementasi Kode Program TF-IDF dan <i>Improved Sqrt-Cosine Similarity</i> .....	124
5.6.3	Implementasi Kode Program Word2Vec dan <i>Cosine Similarity</i> .....	129
5.6.4	Implementasi Kode Program Word2Vec dan <i>Improved Sqrt-Cosine Similarity</i> .....	134
5.7	Implementasi Kode Program Pengujian .....	139
5.7.1	Implementasi Kode Program Perhitungan SRCC .....	139
5.7.2	Implementasi Kode Program Perhitungan Relevansi dan Senioritas.....	144
BAB 6	PENGUJIAN DAN ANALISIS HASIL.....	150
6.1	Pengujian.....	150
6.2	Analisis Hasil.....	156
BAB 7	PENUTUP .....	165
7.1	Kesimpulan.....	165
7.2	Saran.....	166

## DAFTAR TABEL

Tabel 1.1 Tingkat pengangguran 7 negara ASEAN World Economic Outlook .....	1
Tabel 2.1 Hasil eksperimen pertama penelitian oleh Ahmad Alsharef dkk.....	5
Tabel 2.2 Hasil eksperimen kedua penelitian oleh Ahmad Alsharef dkk.....	6
Tabel 4.1 Data uji resume untuk perhitungan manual .....	44
Tabel 4.2 Data uji kualifikasi lowongan kerja untuk perhitungan manual .....	45
Tabel 4.3 Hasil perhitungan manual ekstraksi <i>section</i> .....	45
Tabel 4.4 Hasil perhitungan manual <i>preprocessing</i> resume bagian menghapus <i>email</i> .....	46
Tabel 4.5 Hasil perhitungan manual <i>preprocessing</i> resume bagian menghapus nomor telepon .....	47
Tabel 4.6 Hasil perhitungan manual <i>preprocessing</i> resume bagian menghapus berbagai tipe minus (−, −, −) .....	48
Tabel 4.7 Hasil perhitungan manual <i>preprocessing</i> resume bagian menghapus nama bulan.....	48
Tabel 4.8 Hasil perhitungan manual <i>preprocessing</i> resume bagian menghapus kata “Present” dan “Current” .....	49
Tabel 4.9 Hasil perhitungan manual <i>preprocessing</i> resume bagian menghapus tanggal.....	50
Tabel 4.10 Hasil perhitungan manual <i>preprocessing</i> resume bagian menghapus <i>placeholder</i> .....	51
Tabel 4.11 Hasil perhitungan manual <i>preprocessing</i> resume bagian menghapus tanda baca.....	52
Tabel 4.12 Hasil Perhitungan manual <i>preprocessing</i> resume bagian menghapus angka .....	53
Tabel 4.13 Hasil perhitungan manual <i>preprocessing</i> resume bagian menghapus spasi kosong berlebih.....	54
Tabel 4.14 Hasil perhitungan manual <i>preprocessing</i> resume bagian lematisasi dan menghapus <i>stop words</i> .....	54
Tabel 4.15 Hasil perhitungan manual penyetaraan nama <i>section</i> bagian mengonversi nama <i>section</i> menjadi huruf kecil ( <i>lower casing</i> ) .....	55
Tabel 4.16 Hasil perhitungan manual penyetaraan nama <i>section</i> bagian mengonversi nama <i>section</i> menjadi huruf kecil ( <i>lower casing</i> ) .....	56
Tabel 4.17 hasil perhitungan manual penyetaraan nama <i>section</i> bagian menyeragamkan pengelompokan <i>section</i> berdasarkan pemetaan .....	57

Tabel 4.18 Hasil perhitungan manual penyetaraan nama <i>section</i> bagian klasifikasi, penghapusan, dan pengelompokan <i>section</i> tidak valid.....	58
Tabel 4.19 Hasil perhitungan manual penyetaraan nama <i>section</i> bagian mengonversi isi resume menjadi huruf kecil ( <i>lower casing</i> ) .....	58
Tabel 4.20 Hasil perhitungan manual <i>preprocessing</i> kualifikasi lowongan kerja bagian mengonversi isi kualifikasi lowongan kerja menjadi huruf kecil ( <i>lower casing</i> ) .....	59
Tabel 4.21 Hasil perhitungan manual <i>preprocessing</i> kualifikasi lowongan kerja bagian menghapus angka .....	59
Tabel 4.22 Hasil perhitungan manual <i>preprocessing</i> kualifikasi lowongan kerja bagian menghapus tanda baca .....	60
Tabel 4.23 Hasil perhitungan manual <i>preprocessing</i> kualifikasi lowongan kerja bagian menghapus spasi kosong berlebih .....	60
Tabel 4.24 Hasil perhitungan manual <i>preprocessing</i> kualifikasi lowongan kerja bagian lematisasi dan menghapus <i>stop words</i> .....	61
Tabel 4.25 Korpus resume untuk perhitungan manual .....	61
Tabel 4.26 Perhitungan manual frekuensi <i>term</i> setiap resume .....	62
Tabel 4.27 Perhitungan manual TF korpus resume .....	64
Tabel 4.28 Perhitungan manual IDF korpus resume.....	66
Tabel 4.29 Perhitungan manual TF-IDF korpus resume.....	68
Tabel 4.30 Perhitungan manual TF-IDF korpus resume setelah normalisasi .....	71
Tabel 4.31 Korpus kualifikasi lowongan kerja untuk perhitungan manual.....	73
Tabel 4.32 Perhitungan manual TF korpus kualifikasi lowongan kerja .....	73
Tabel 4.33 Perhitungan manual IDF korpus kualifikasi lowongan kerja .....	76
Tabel 4.34 Perhitungan manual TF-IDF korpus kualifikasi lowongan kerja.....	78
Tabel 4.35 Perhitungan manual TF-IDF korpus kualifikasi lowongan kerja setelah normalisasi .....	80
Tabel 4.36 Perhitungan manual <i>one-hot encoding</i> .....	83
Tabel 4.37 Bobot <i>input layer-hidden layer</i> .....	85
Tabel 4.38 Bobot <i>hidden layer-output layer</i> .....	86
Tabel 4.39 Bobot <i>hidden layer-output layer</i> .....	86
Tabel 4.40 Bobot <i>hidden layer-output layer</i> .....	86
Tabel 4.41 Bobot <i>hidden layer-output layer</i> .....	86
Tabel 4.42 Pembaharuan bobot <i>input layer-hidden layer</i> .....	91

Tabel 4.43 Pembaharuan bobot <i>input layer-output layer</i> .....	92
Tabel 4.44 Pembaharuan bobot <i>input layer-output layer</i> .....	93
Tabel 4.45 Pembaharuan bobot <i>input layer-output layer</i> .....	93
Tabel 4.46 Pembaharuan bobot <i>input layer-output layer</i> .....	93
Tabel 4.47 Data sampel perhitungan manual similaritas .....	94
Tabel 4.48 Vektor TF-IDF perhitungan manual <i>Improved Sqrt-Cosine Similarity</i> .	95
Tabel 4.49 Hasil skor similaritas resume ID 15265464 perhitungan manual <i>Improved Sqrt-Cosine Similarity</i> .....	96
Tabel 4.50 Bobot <i>section</i> kategori industri "TEACHER" .....	96
Tabel 4.51 Vektor Word2Vec perhitungan manual <i>Cosine Similarity</i> .....	97
Tabel 4.52 Hasil skor similaritas resume ID 15265464 perhitungan manual <i>Cosine Similarity</i> .....	98
Tabel 4.53 Vektor Word2Vec perhitungan manual <i>Improved Sqrt-Cosine Similarity</i> .....	99
Tabel 4.54 Hasil skor similaritas resume ID 15265464 perhitungan manual <i>Improved Sqrt-Cosine Similarity</i> .....	100
Tabel 4.55 Peringkat 1-5 resume dengan skor similaritas terbesar untuk perhitungan manual SRCC.....	102
Tabel 4.56 Selisih peringkat 1-5 resume perhitungan manual SRCC.....	102
Tabel 4.57 Peringkat 1-5 resume dengan hasil evaluasi relevansi dan senioritas ahli terbesar untuk perhitungan manual.....	103
Tabel 6.1 Hasil pengujian berwarna hijau skenario tanpa bobot <i>section</i> .....	156
Tabel 6.2 Hasil pengujian berwarna merah skenario tanpa bobot <i>section</i> .....	157
Tabel 6.3 Hasil pengujian berwarna hijau skenario dengan bobot <i>section</i> .....	158
Tabel 6.4 Hasil pengujian berwarna merah skenario dengan bobot <i>section</i> .....	158
Tabel 6.5 <i>Weighted score</i> keseluruhan pendekatan dan skenario.....	159
Tabel 6.6 Urutan pendekatan berdasarkan <i>weighted score</i> tertinggi .....	159
Tabel 6.7 Perhitungan rata-rata parameter setiap pendekatan dan skenario ...	160
Tabel 6.8 Perhitungan similaritas antar <i>term</i> Word2Vec vektor nilai asli dengan vektor nilai absolut.....	161

## DAFTAR GAMBAR

Gambar 2.1 Pengelompokan <i>text similarity measure</i> .....	8
Gambar 2.2 Arsitektur pendekatan <i>Skip-gram</i> .....	12
Gambar 3.1 Struktur proses implementasi pemeringkatan similaritas resume dan kualifikasi lowongan kerja.....	19
Gambar 4.1 Diagram alur ekstraksi <i>section</i> .....	21
Gambar 4.2 Diagram alur ekstraksi <i>section</i> .....	22
Gambar 4.3 Diagram alur ekstraksi <i>section</i> .....	23
Gambar 4.4 Diagram alur ekstraksi <i>section</i> .....	24
Gambar 4.5 Diagram alur <i>preprocessing</i> isian resume .....	25
Gambar 4.6 Diagram alur <i>preprocessing</i> isian resume .....	26
Gambar 4.7 Diagram alur <i>preprocessing</i> penamaan <i>section</i> resume.....	28
Gambar 4.8 Diagram alur <i>preprocessing</i> isian kualifikasi lowongan kerja .....	29
Gambar 4.9 Diagram alur TF-IDF.....	30
Gambar 4.10 Diagram alur TF-IDF.....	31
Gambar 4.11 Diagram alur TF-IDF.....	32
Gambar 4.12 Diagram alur Word2Vec .....	33
Gambar 4.13 Diagram alur Word2Vec .....	34
Gambar 4.14 Diagram alur Word2Vec .....	35
Gambar 4.15 Diagram alur <i>Improved Sqrt-Cosine Similarity</i> .....	36
Gambar 4.16 Diagram alur <i>Improved Sqrt-Cosine Similarity</i> .....	37
Gambar 4.17 Diagram alur <i>Cosine Similarity</i> .....	38
Gambar 4.18 Diagram alur <i>Cosine Similarity</i> .....	39
Gambar 4.19 Diagram alur <i>Spearman Rank Correlation Coefficient</i> .....	40
Gambar 4.20 Diagram alur <i>Spearman Rank Correlation Coefficient</i> .....	41
Gambar 4.21 Diagram alur relevansi dan senioritas.....	42
<b>Gambar 4.22 Diagram alur relevansi dan senioritas.....</b>	<b>43</b>
Gambar 4.23 Pasangan target-konteks Word2Vec <i>Skip-gram</i> .....	83
Gambar 6.1 Cuplikan <i>template spreadsheet</i> evaluasi ahli.....	150
Gambar 6.2 Cuplikan <i>template spreadsheet</i> evaluasi ahli.....	151
Gambar 6.3 Cuplikan <i>spreadsheet</i> hasil evaluasi ahli .....	151

Gambar 6.4 Hasil <i>descriptive statistics</i> SPSS.....	152
Gambar 6.5 Visualisasi nilai korelasi .....	153
Gambar 6.6 Visualisasi persentase relevansi.....	154
Gambar 6.7 Visualisasi persentase senioritas.....	155
Gambar 6.8 Visualisasi pergeseran posisi <i>term</i> Word2Vec vektor nilai asli dengan vektor nilai absolut.....	161
Gambar 6.9 Visualisasi pergeseran posisi <i>term</i> Word2Vec vektor nilai asli dengan vektor nilai absolut.....	162
Gambar 6.10 Grafik garis tiga parameter setiap kualifikasi lowongan kerja.....	163

## **DAFTAR LAMPIRAN**

LAMPIRAN A SURAT PERNYATAAN VALIDITAS .....	171
LAMPIRAN B BOBOT PER SECTION BERDASARKAN INDUSTRI.....	172
LAMPIRAN C HASIL PEMERINGKATAN LIMA RESUME PER KUALIFIKASI LOWONGAN KERJA .....	180
C.1 Tanpa Bobot - TF-IDF dan <i>Improved Sqrt-Cosine Similarity</i> .....	180
C.2 Tanpa Bobot - Word2Vec dan <i>Cosine Similarity</i> .....	190
C.3 Tanpa Bobot - Word2Vec dan <i>Improved Sqrt-Cosine Similarity</i> .....	201
C.4 Dengan Bobot - TF-IDF dan <i>Improved Sqrt-Cosine Similarity</i> .....	211
C.5 Dengan Bobot - Word2Vec dan <i>Cosine Similarity</i> .....	220
C.6 Dengan Bobot - Word2Vec dan <i>Improved Sqrt-Cosine Similarity</i> .....	231
LAMPIRAN D GRAFIK GARIS TIGA PARAMETER SETIAP KUALIFIKASI LOWONGAN KERJA .....	242

# BAB 1 PENDAHULUAN

Bab pendahuluan membahas mengenai latar belakang penelitian, rumusan masalah, tujuan, manfaat, batasan masalah, serta sistematika pembahasan dari penelitian ini.

## 1.1 Latar Belakang

Tingkat pengangguran yang tinggi merupakan salah satu tantangan utama yang dihadapi Indonesia. Masalah ini tidak hanya mempengaruhi kondisi perekonomian, tetapi juga kesejahteraan sosial masyarakat. Perkembangan ketenagakerjaan sangat penting bagi stabilitas ekonomi, dan setiap hambatan dalam aspek ini dapat berdampak negatif pada upaya meningkatkan taraf hidup masyarakat.

**Tabel 1.1 Tingkat pengangguran 7 negara ASEAN World Economic Outlook**

No	Negara	Tingkat Pengangguran
1	Indonesia	5,2
2	Filipina	5,1
3	Brunei Darussalam	4,9
4	Malaysia	3,5
5	Viet Nam	2,1
6	Singapore	1,9
7	Thailand	1,1

Tabel 1.1 menunjukkan data yang diambil dari World Economic Outlook pada April 2024 oleh International Monetary Fund, di antara 7 negara ASEAN yang datanya tercantum, Indonesia memiliki tingkat pengangguran tertinggi di angka 5,2. Pengangguran bisa disebabkan oleh beberapa fenomena dan salah satunya adalah ketidaksesuaian antara karakteristik pencari kerja dengan tawaran kerja atau bisa disebut dengan pengangguran struktural (Kementerian Ketenagakerjaan RI - Badan Perencanaan dan Pengembangan Ketenagakerjaan, 2021). Sumber daya, kualifikasi, keterampilan, dan pengetahuan yang tersedia dan diperoleh oleh individu untuk memaksimalkan kemampuan kerja mereka sendiri disebut dengan *human capital*. Nilai-nilai ini berkontribusi pada pendapatan yang lebih tinggi, kepuasan hidup, dan kohesi sosial, sehingga juga menjadi salah satu penentu pertumbuhan ekonomi negara (Wujarso, 2022).

Proses mencari dan melamar pekerjaan identik dengan penggunaan *curriculum vitae* (CV) atau beberapa orang menyebutnya resume. Tahapan awal perekrutan adalah proses *screening* CV yang selanjutnya diikuti dengan proses wawancara. Rekrutmen dianggap efektif ketika mendapatkan banyak pelamar yang sesuai dengan kualifikasi untuk mendapatkan calon karyawan terbaik dari yang terbaik (Budiantoro dalam Kumaladewi, 2018).

Terdapat beberapa penelitian sebelumnya yang berkaitan dengan perhitungan similaritas teks untuk otomatisasi penyaringan resume. Pertama, penelitian oleh Ahmad Alsharef dkk. (2023) membandingkan pendekatan *Cosine Similarity*, *Sqrt-Cosine Similarity*, dan *Improved Sqrt-Cosine (ISC) Similarity*, menggunakan TF-IDF sebagai metode vektorisasi teks, dan menunjukkan bahwa pendekatan ISC dianggap lebih baik dibandingkan dua pendekatan lainnya. Kedua, penelitian oleh Rahul Singh Pundir dkk. (2024) mengembangkan sistem rekomendasi resume berbasis keterampilan, menggunakan Word2Vec untuk menangkap kesamaan semantik keterampilan dan *Cosine Similarity* untuk mengukur kesesuaian antara vektor keterampilan dengan kebutuhan pekerjaan. Merujuk pada penelitian sebelumnya, penelitian ini bertujuan untuk mengimplementasikan *Improved Sqrt-Cosine (ISC) Similarity* dalam memeringkat resume berdasarkan kualifikasi lowongan kerja, dengan mengeksplorasi metode representasi teks TF-IDF dan Word2Vec.

## 1.2 Rumusan Masalah

Berikut ini merupakan rumusan masalah penelitian.

1. Bagaimana hasil pemeringkatan lima resume untuk setiap kualifikasi lowongan kerja menggunakan *Improved Sqrt-Cosine Similarity* dalam mengkalkulasikan similaritas teks?
2. Bagaimana korelasi antara peringkat hasil implementasi *Improved Sqrt-Cosine Similarity* dengan peringkat hasil evaluasi ahli untuk setiap kualifikasi lowongan kerja menggunakan *Spearman Rank Correlation Coefficient*?

## 1.3 Tujuan

Berikut ini merupakan tujuan penelitian ini.

1. Menganalisis hasil pemeringkatan lima resume untuk setiap kualifikasi lowongan kerja menggunakan *Improved Sqrt-Cosine Similarity* dalam mengkalkulasikan similaritas teks.
2. Menganalisis korelasi antara peringkat hasil implementasi *Improved Sqrt-Cosine Similarity* dengan peringkat hasil evaluasi ahli untuk setiap kualifikasi lowongan kerja menggunakan *Spearman Rank Correlation Coefficient*.

## **1.4 Manfaat**

Berikut merupakan manfaat yang dapat diperoleh dari penelitian ini.

1. Memberikan interpretasi hasil *Improved Sqrt-Cosine Similarity* dalam menghasilkan perhitungan similaritas teks pada pemeringkatan lima resume untuk setiap kualifikasi lowongan kerja.
2. Memberikan pemaparan mengenai korelasi antara peringkat hasil *Improved Sqrt-Cosine Similarity* dengan peringkat hasil evaluasi ahli menggunakan *Spearman Rank Correlation Coefficient*.

## **1.5 Batasan Masalah**

Batasan masalah yang ditetapkan dalam penelitian ini sebagai berikut.

1. Penelitian ini terbatas pada *dataset* Kaggle dengan 2.484 resume.
2. Penelitian ini berfokus pada pemeringkatan lima resume berdasarkan nilai similaritas tertinggi untuk setiap kualifikasi lowongan kerja.
3. Penelitian ini melibatkan seorang ahli di bidang rekrutmen dalam mengevaluasi hasil pemeringkatan.
4. Kualifikasi lowongan kerja yang digunakan diambil dari 24 posisi di portal lowongan pekerjaan <https://id.jobstreet.com/>.

## **1.6 Sistematika Pembahasan**

Susunan sistematika pembahasan ditulis di bawah ini dan terdiri dari beberapa bab yang menjelaskan mengenai penelitian mengenai kalkulasi similaritas teks pada resume pelamar dengan kualifikasi lowongan kerja.

### **BAB 1 PENDAHULUAN**

Bab pendahuluan membahas mengenai latar belakang penelitian, rumusan masalah, tujuan, manfaat, batasan masalah, serta sistematika pembahasan dari penelitian ini.

### **BAB 2 LANDASAN KEPUSTAKAAN**

Bab landasan kepustakaan berisi kajian pustaka dan dasar teori. Penelitian-penelitian sebelumnya yang berhubungan dengan kalkulasi similaritas teks pada resume dijelaskan pada kajian pustaka. Sedangkan, penjelasan teori, konsep, dan metode yang digunakan dijelaskan pada dasar teori.

### **BAB 3 METODOLOGI PENELITIAN**

Terdapat beberapa bagian di bab metodologi penelitian, seperti tipe penelitian, strategi penelitian, lokasi penelitian, metode pengumpulan data, metode analisis data, peralatan pendukung, dan perancangan algoritma.

### **BAB 4 PERANCANGAN**

Bab perancangan menjelaskan tentang perancangan algoritma dari metode-metode serta *flow diagram* dari setiap algoritma yang digunakan pada penelitian ini. Selain itu, bab ini juga merincikan perhitungan manual.

## **BAB 5 IMPLEMENTASI**

Bab implementasi berisi implementasi dari metode kalkulasi similaritas teks pada resume pelamar dengan kualifikasi instansi yang digunakan pada penelitian ini, seperti metode perhitungan similaritas *Improved Sqrt-Cosine* (ISC) dan *Cosine Similarity* (CosSim), serta metode representasi teks TF-IDF dan Word2Vec.

## **BAB 6 PENGUJIAN DAN ANALISIS HASIL**

Pemaparan hasil dari pengujian akan dijelaskan di bab pengujian, serta pembahasan dan analisa dari hasil pengujian tersebut sebagai bahan evaluasi.

## **BAB 7 PENUTUP**

Terakhir, kesimpulan dan saran ditulis pada bab penutup. Bagian kesimpulan memaparkan rangkuman dari hasil penelitian untuk menjawab semua rumusan masalah yang dijabarkan pada latar belakang. Sedangkan, bagian saran memaparkan masukan-masukan untuk penelitian selanjutnya agar penelitian ini dapat diperbaiki dan dikembangkan.

## BAB 2 LANDASAN KEPUSTAKAAN

Bab landasan kepustakaan berisi kajian pustaka dan dasar teori. Penelitian-penelitian sebelumnya yang berhubungan dengan kalkulasi similaritas teks pada resume dijelaskan pada kajian pustaka. Sedangkan, penjelasan teori, konsep, dan metode yang digunakan dijelaskan pada dasar teori.

### 2.1 Kajian Pustaka

Terdapat beberapa penelitian yang dilakukan sebelumnya terkait perhitungan similaritas teks untuk otomatisasi penyaringan resume. Penelitian pertama dilakukan oleh Ahmad Alsharef dkk. (2023) berjudul "*Exploring the Efficiency of Text-Similarity Measures in Automated Resume Screening for Recruitment*" mengeksplorasi penggunaan *text similarity* sebagai alternatif dalam memproses resume, dengan pendekatan *Cosine Similarity*, *Sqrt-Cosine Similarity*, dan *Improved Sqrt-Cosine (ISC) Similarity*. Terdapat dua eksperimen dalam penelitian ini. Eksperimen pertama melibatkan 40 resume yang disandingkan dengan deskripsi pekerjaan untuk posisi manajer pengembangan bisnis di salah satu *platform* terkemuka untuk manajemen kepatuhan UKM di Eropa. Pada lima urutan teratas peringkat resume, dilakukan perbandingan antara peringkat yang diberikan oleh manusia dan pengukuran otomatis, hasilnya tertera pada Tabel 2.1

**Tabel 2.1 Hasil eksperimen pertama penelitian oleh Ahmad Alsharef dkk.**

<i>Human Ranking of Resumes</i>	<i>ISC</i>	<i>Sqrt-Cosine</i>	<i>Cosine</i>
1	2 (0,38)	11 (28,545)	7 (43,02)
2	3 (0,345)	2 (32,732)	1 (48,44)
3	1 (0,391)	9 (28,87)	4 (44,29)
4	4 (0,344)	17 (26,749)	14 (38,5)
5	8 (0,302)	1 (32,733)	15 (38,2)

Eksperimen kedua melibatkan 30 resume yang disandingkan dengan deskripsi pekerjaan untuk posisi *software engineer* di salah satu perusahaan teknologi multinasional Amerika. Pada lima urutan teratas peringkat resume, dilakukan perbandingan antara peringkat yang diberikan oleh manusia dan pengukuran otomatis, yang hasilnya tertera pada Tabel 2.2.

**Tabel 2.2 Hasil eksperimen kedua penelitian oleh Ahmad Alsharef dkk.**

<i>Human Ranking of Resumes</i>	<i>ISC</i>	<i>Sqrt-Cosine</i>	<i>Cosine</i>
1	8 (0,181)	6 (21,205)	11 (18,47)
2	1 (0,246)	4 (22,136)	1 (26,99)
3	4 (0,199)	1 (22,744)	4 (22,25)
4	3 (0,206)	2 (22,65)	10 (18,52)
5	6 (0,181)	14 (18,708)	3 (24,36)

Berdasarkan hasil kedua eksperimen, hasil dari penelitian yang dilakukan oleh Ahmad Alsharef dkk. (2023) menunjukkan bahwa peringkat *Improved Sqrt-Cosine* (ISC) *Similarity* cenderung lebih mendekati peringkat yang diberikan oleh manusia.

Penelitian kedua dilakukan oleh Rahul Singh Pundir dkk. (2024) berjudul "*Enhancing Resume Recommendation System through Skill-based Similarity using Deep Learning Models*" membahas cara meningkatkan rekomendasi resume dengan mempertimbangkan kesamaan keterampilan. Sistem ini menggunakan metode Word2Vec untuk mengukur kecocokan kandidat dengan kebutuhan pekerjaan berdasarkan *skills* dan LSTM-RNN untuk memprediksi profil pekerjaan. Skor *skill similarity* dari penelitian ini berkisar dari 0,447 hingga 0,790, di mana nilai yang lebih tinggi menunjukkan keterampilan kandidat lebih sesuai dengan kebutuhan pekerjaan. Pendekatan ini juga membantu kandidat memahami keterampilan tambahan yang perlu dikembangkan untuk memenuhi kualifikasi posisi yang diinginkan.

Penelitian ketiga dilakukan oleh Chirag Daryani dkk. (2020) berjudul "*An Automated Resume Screening System Using Natural Language Processing and Similarity*" mengembangkan sistem rekomendasi yang mengekstraksi informasi dari resume yang tidak terstruktur dan mengubahnya menjadi vektor yang mewakili fitur-fitur penting seperti pendidikan, pengalaman, dan keterampilan. Dengan menghitung kesamaan menggunakan *Cosine Similarity* antara resume dan deskripsi pekerjaan, sistem ini mampu menyusun peringkat kandidat terbaik yang sesuai dengan posisi pekerjaan yang ditawarkan. Hasil perhitungan *Cosine Similarity* antara empat resume dan *query* pekerjaan menunjukkan bahwa resume kandidat ke-2 menduduki peringkat pertama (0,680), diikuti resume kandidat ke-4 di peringkat kedua (0,651), resume kandidat ke-3 di peringkat ketiga (0,498), dan resume kandidat ke-1 di peringkat terakhir (0,490).

Topik skripsi yang diambil memiliki beberapa kesamaan dengan penelitian sebelumnya. Pertama, penelitian ini mengimplementasikan perhitungan similaritas *Improved Sqrt-Cosine* (ISC), seperti pada penelitian pertama. Kedua, cara menghitung similaritas dilakukan dengan membandingkan informasi yang didapatkan dari resume dengan persyaratan atau kualifikasi posisi pekerjaan, sebagaimana dilakukan pada penelitian pertama dan ketiga. Ketiga, penelitian ini mengekstaksi informasi dari resume seperti pada penelitian ketiga, tetapi berfokus pada pengambilan informasi per bagian (*section*) resume sesuai standarisasi dari seorang ahli. Perbedaan dengan penelitian kedua adalah penelitian tersebut menggunakan daftar *skills* yang diambil dari kumpulan resume dan mengukur *skill similarity*-nya dengan resume yang digunakan. Sedangkan, topik skripsi ini akan mengukur *similarity* dari suatu resume untuk melihat apakah sesuai dengan yang dibutuhkan oleh suatu kualifikasi posisi pekerjaan. Meskipun begitu, untuk tahap representasi teks juga akan menggunakan Word2Vec seperti yang dilakukan pada penelitian kedua sebagai salah satu metode representasi teks dari penelitian ini.

## 2.2 Dasar Teori

### 2.2.1 Resume

Resume menurut Kamus Besar Bahasa Indonesia (KBBI) merupakan kata nominal yang berarti ikhtisar atau ringkasan. Stanford Career Education (2024) menyatakan bahwa resume merupakan ringkasan pengalaman yang dipilih oleh perekrut untuk menunjukkan kesesuaian pelamar dengan posisi yang dilamar. Resume juga sering diartikan sama dengan *Curriculum Vitae* (CV), keduanya pun memiliki definisi yang sama menurut Cambridge Dictionary (2024), yaitu sebuah ringkasan tertulis yang menggambarkan latar belakang pendidikan, kualifikasi, pengalaman kerja sebelumnya, serta minat pribadi seseorang dan dikirimkan kepada instansi ketika melamar pekerjaan. Di Amerika Serikat, CV umumnya digunakan saat melamar pekerjaan di bidang akademis, sedangkan resume digunakan untuk pekerjaan lainnya (Cambridge University Press & Assessment, 2024).

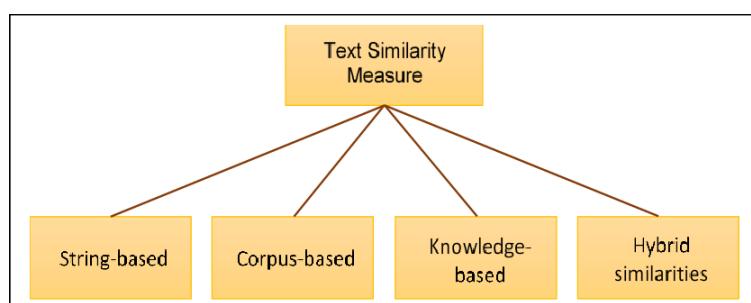
Perekrut hanya meluangkan kurang dari 30 detik untuk meninjau resume, sehingga penting bagi pelamar untuk secara cepat dan jelas menunjukkan bagaimana pelatihan dan pengalaman mereka dapat memberi nilai tambah bagi perusahaan, serta memaparkan keterampilan secara tepat untuk memenuhi kebutuhan perusahaan dengan format resume yang jelas, menarik, dan disesuaikan dengan konvensi yang berlaku di setiap posisi atau sektor yang dilamar (Stanford Career Education, 2018).

Stanford Career Education (2024) juga memberikan panduan dalam langkah-langkah membuat resume, dikatakan bahwa resume yang efektif adalah ringkasan singkat yang menyoroti pengalaman dan keterampilan yang langsung terkait dengan pekerjaan. Informasi yang ingin disampaikan kepada perekrut mengenai masing-masing pengalaman harus ditentukan dengan jelas, konten yang difokuskan tergantung pada posisi yang dilamar karena setiap perekrut mencari sekumpulan keterampilan tertentu dari pelamar yang sesuai dengan keterampilan yang diperlukan untuk menjalankan pekerjaan tertentu (Stanford Career Education, 2024). Saat mendeskripsikan keterampilan atau pencapaian yang relevan, disarankan untuk menggunakan metode C-A-R: *CONTEXT* mencakup apa yang dikerjakan, seperti tugas, proyek, atau tujuan keseluruhan yang tercapai, serta pihak-pihak yang terlibat, seperti tim yang berkolaborasi dan/atau populasi yang dilayani; *ACTIONS* menggambarkan bagaimana tugas tersebut dilaksanakan dengan menekankan keterampilan yang digunakan untuk menyelesaikan tugas, penggunaan *action words* sangat dianjurkan untuk mendeskripsikan tindakan yang diambil; Terakhir, *RESULTS* menjelaskan apa hasilnya, kuantifikasi hasil harus dilakukan jika memungkinkan, atau jika hasil tidak diketahui, penting untuk menyertakan tujuan dari tugas, proyek, atau tujuan tersebut, yang menjelaskan alasan pelaksanaan (Stanford Career Education, 2024).

### 2.2.2 Similaritas Teks

Mengingat tujuan perekrut adalah mencari pelamar yang memiliki sekumpulan keterampilan yang dibutuhkan oleh instansi untuk melaksanakan pekerjaan tertentu, maka digunakan metode untuk menghitung similaritas antara kualifikasi yang terdapat dalam resume pelamar dengan kualifikasi yang dibutuhkan oleh instansi. Similaritas teks adalah membandingkan suatu teks dengan teks lainnya dan menemukan persamaan di antara mereka. Pada dasarnya, ini tentang menentukan tingkat kedekatan teks tersebut. Dalam pemrosesan bahasa alami, menentukan apakah makna dari dua dokumen identik adalah tugas mendasar dan luas yang memungkinkan komputer memahami bahasa manusia (He, et al., 2024).

Berbagai pendekatan telah dikembangkan untuk mengukur similaritas antara satu teks dengan teks lainnya yang terbagi menjadi empat kelompok utama, yaitu *String-based*, *Corpus-based*, *Knowledge-based*, dan *Hybrid* (Prasetya, et al., 2018).



Gambar 2.1 Pengelompokan *text similarity measure*

Sumber: Prasetya et al. (2018)

Seperti pada apa yang sudah diilustrasikan pada Gambar 2.1, *String-based* adalah metode pengukuran tertua, paling sederhana, tetapi paling populer dan beroperasi pada urutan string dan susunan karakter (Prasetya, et al., 2018). *Corpus-based* menggunakan pendekatan semantik yang mana menentukan kesamaan antara dua konsep berdasarkan informasi yang diekstraksi dari korpus yang besar (Prasetya, et al., 2018). *Knowledge-based* menggunakan hubungan semantik untuk mengidentifikasi tingkat kesamaan kata-kata (Prasetya, et al., 2018). Selain tiga kelompok yang telah dijelaskan sebelumnya, terdapat perhitungan similaritas secara *Hybrid* dengan tujuan untuk menggabungkan metode yang telah disebutkan sebelumnya, termasuk *String-based*, *Corpus-based*, dan *Knowledge-based* guna mencapai metrik yang lebih baik dengan mengadopsi keunggulan masing-masing metode (Prasetya, et al., 2018).

### 2.2.3 Pra-pemrosesan Teks

Sebelum menerapkan metode similaritas teks, diperlukan proses pra-pemrosesan teks terlebih dahulu untuk menyiapkan *dataset* resume yang akan diolah. Penting untuk melakukan pemrosesan data ini guna memastikan bahwa data yang digunakan adalah data yang berkualitas sehingga dapat menghasilkan akurasi yang akurat (Prasetya, et al., 2024). Adapun beberapa langkah selama proses pra-pemrosesan teks, yaitu:

1. *Tokenization*. Proses ini untuk memecahkan kalimat menjadi kata-kata (Sohangir & Wang, 2017). *Tools* seperti NLTK dan spaCy biasa digunakan untuk melakukan tokenisasi (Amin, et al., 2023).
2. *Lower Casing*. Proses ini mengonversi semua teks menjadi huruf kecil (Alsharef, et al., 2023).
3. *Stop Words*. Proses ini untuk menghilangkan *stop words*, yaitu kata-kata umum dalam dokumen yang tidak memiliki makna signifikan dan tidak berkontribusi dalam membedakan dokumen, sehingga dapat diabaikan (Sihombing, 2022). Bahasa pemrograman Python telah menyediakan *package Regular Expression* (RegEx) untuk melaksanakan penghilangan *stop words* (Amin, et al., 2023).
4. *Lemmatization*. Tidak seperti *stemming* yang hanya mengubah kata menjadi ke bentuk dasarnya, *lemmatization* memanfaatkan kosakata dan morfologis yang sesuai dengan linguistik (Daryani, et al., 2020) sesuai kamus bahasa yang digunakan. Untuk bahasa Inggris, dapat memanfaatkan WordNet Lemmatizer yang tersedia melalui NLTK Python (Daryani, et al., 2020).

## 2.2.4 TF-IDF

*Term Frequency-Inverse Document Frequency* (TF-IDF) adalah salah satu teknik yang digunakan untuk menghitung seberapa penting suatu kata (*term*) dalam sebuah dokumen terhadap keseluruhan kumpulan dokumen (Septiani & Isabela, 2022). TF-IDF didapatkan dari hasil perkalian antara *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF) (Ramadhan, et al., 2023).

$$TF - IDF_{(t,d)} = TF_{(t,d)} \times IDF_{(t)} \quad (2.1)$$

Adapun keterangan dari Persamaan 2.1:

- $TF - IDF_{(t,d)}$  = Bobot TF-IDF pada *term* ke-*t* terhadap dokumen ke-*d*  
 $TF_{(t,d)}$  = Frekuensi kemunculan *term* ke-*t* terhadap dokumen ke-*d*  
 $IDF_{(t)}$  = Nilai *inverse* dokumen yang memunculkan *term* ke-*t*

TF didapatkan dengan menghitung jumlah kemunculan kata dibagi dengan total kata dalam dokumen (Septiani & Isabela, 2022). IDF menghitung seberapa penting suatu kata dalam koleksi dokumen dengan membagi jumlah total dokumen dengan jumlah dokumen yang mengandung suatu *term* (Septiani & Isabela, 2022).

$$TF_{(t,d)} = \frac{f_{t,d}}{\sum_k f_{k,d}} \quad (2.2)$$

Adapun keterangan dari Persamaan 2.2:

- $TF_{(t,d)}$  = Frekuensi kemunculan *term* ke-*t* terhadap dokumen ke-*d*  
 $f_{t,d}$  = Jumlah kemunculan *term* ke-*t* terhadap dokumen ke-*d*  
 $\sum_k f_{k,d}$  = Total seluruh *term* terhadap dokumen ke-*d*

$$IDF_{(t)} = \log \left( \frac{N}{df} \right) \quad (2.3)$$

Keterangan dari Persamaan 2.3:

- $IDF_{(t)}$  = Nilai *inverse document frequency* untuk *term* ke-*t*  
 $N$  = Jumlah total dokumen  
 $df$  = Banyaknya dokumen yang mengandung *term*

Dalam dokumentasi Scikit-learn, TF *default* tidak dinormalisasi dengan membagi total *term*. Normalisasi baru dilakukan setelah mendapatkan hasil perkalian TF dengan IDF menggunakan  $L_2$  *normalization*, sehingga TF didapatkan dari jumlah kemunculan (frekuensi) *term* terhadap suatu dokumen seperti pada Persamaan 2.4

$$TF_{(t,d)} = f_{t,d} \quad (2.4)$$

Keterangan:

$TF_{(t,d)}$  = Frekuensi kemunculan *term* ke-*t* terhadap dokumen ke-*d*

$f_{t,d}$  = Jumlah kemunculan *term* ke-*t* terhadap dokumen ke-*d*

Pada dokumentasi Scikit-learn, IDF *default* menggunakan mekanisme *smoothing* dengan menambahkan konstanta "1" pada pembilang dan penyebut, seolah-olah ada dokumen tambahan yang mengandung setiap *term* dalam koleksi tepat satu kali sehingga mencegah pembagian oleh nol. Rumusnya menjadi seperti pada Persamaan 2.5

$$IDF_{(t)} = \log\left(\frac{1+N}{1+df}\right) + 1 \quad (2.5)$$

Keterangan:

$IDF_{(t)}$  = Nilai *inverse document frequency* untuk *term* ke-*t*

$N$  = Jumlah total dokumen

$df$  = Banyaknya dokumen yang mengandung *term*

Setelah mengalikan TF dan IDF, perhitungan TF-IDF dalam dokumentasi Scikit-learn dinormalisasikan dengan  $L_2$  *normalization* atau *Euclidean norm*, formula dari normalisasi ini dapat tertera pada Persamaan 2.6

$$\nu_{norm} = \frac{\nu}{\|\nu\|_2} = \frac{\nu}{\sqrt{\nu_1^2 + \nu_2^2 + \dots + \nu_n^2}} \quad (2.6)$$

Keterangan:

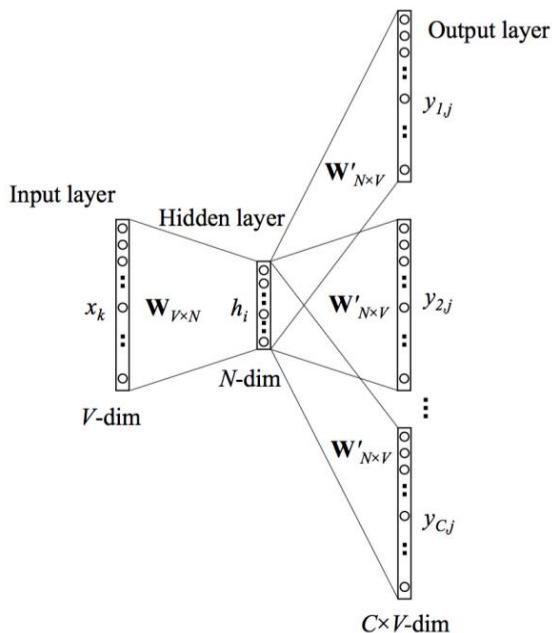
$\nu_{norm}$  = Vektor yang telah dinormalisasi

$\nu$  = Vektor asli sebelum dinormalisasikan

$\|\nu\|_2$  = Akar kuadrat dari jumlah kuadrat semua elemen vektor  $\nu$

## 2.2.5 Word2Vec

Word2Vec adalah metode yang digunakan untuk menghasilkan *word embedding* dengan memanfaatkan *neural networks* sederhana yang dilatih untuk memahami konteks linguistik kata. Pendekatan ini menggunakan *continuously sliding Skip-gram* atau *continuously sliding Bag-of-Words* (CBOW). Word2Vec mengonversi kata-kata menjadi vektor, memungkinkan pengenalan hubungan semantik dan telah menjadi kunci dalam perkembangan berbagai aplikasi *Natural Language Processing* (NLP) (Kulshretha & Lodha, 2023). Pada pendekatan CBOW, suatu kata diprediksi berdasarkan konteks yang mengelilinginya di dalam sebuah kalimat. Sedangkan, pendekatan *Skip-gram* memprediksi konteks berdasarkan kata yang diberikan (Meyer, 2016). Mengingat tujuan utama dari penelitian ini adalah memperhitungkan similaritas teks, sehingga akan lebih fokus pada penggunaan pendekatan *Skip-gram* yang arsitekturnya terlampir pada Gambar 2.2



**Gambar 2.2 Arsitektur pendekatan *Skip-gram***

Sumber: Meyer (2016)

*Skip-gram* bekerja dengan menggunakan kata yang sedang diproses (*current word*) sebagai *input* untuk mempelajari dan memprediksi kata-kata dalam konteks sebagai target. Proses ini mempelajari distribusi probabilitas kata-kata dalam sebuah kalimat berdasarkan jarak antara kata *input* dan kata-kata konteks (*windows*) (Ayuningtyas & Tantyoko, 2024).

Dalam teknik *Skip-gram*, proses *training* dan *inference* dilakukan secara terpisah. Selama proses *training*, *skip-gram* mempelajari konteks dari kata-kata yang muncul di sekitar kata target dalam *window* tertentu. Sebagai contoh, jika ukuran *window* adalah dua, maka kata-kata seperti 'Saya,' 'suka,' 'makan,' dan 'apel' menjadi konteks bagi kata 'apel' dalam kalimat 'Saya suka makan apel.' Metode ini digunakan untuk menghasilkan distribusi probabilitas dari semua kemungkinan konteks kata berdasarkan kata target (Dwivedi & Anand, 2023).

Pada dokumentasi Gensim, implementasinya menggunakan beberapa parameter seperti *sg* yang ditetapkan dengan nilai 1 untuk memakai *Skip-gram*, *vector\_size* untuk menetapkan dimensi vektor-vektor kata, *window* untuk menetapkan jarak maksimum antara kata saat ini dan kata yang diprediksi dalam sebuah kalimat, *alpha* untuk menginisialisasi *learning rate*, dan *epochs* untuk menetapkan iterasi yang secara *default* bernilai lima.

### **2.2.6 Improved Sqrt-Cosine Similarity**

Sohangir dan Wang (2017) memperkenalkan sebuah teknik pengukuran *similarity* yang disebut *Improved Sqrt-Cosine* (ISC) *similarity*, yang didasarkan pada normalisasi  $L_1$  (*Hellinger distance*) dan telah terbukti bahwa pada data berdimensi tinggi, normalisasi  $L_1$  bekerja lebih baik daripada normalisasi  $L_2$  (*Euclidean distance*). Pada persamaan ISC, alih-alih menggunakan normalisasi  $L_1$ , digunakan akar kuadrat dari normalisasi  $L_1$  (Sohangir & Wang, 2017). Sebagian besar menganggap *Cosine Similarity* sebagai 'state of the art' dalam pengukuran *similarity* (Sohangir & Wang, 2017). Melalui eksperimen yang mendalam, diamati bahwa meskipun ISC mirip dengan *Cosine Similarity* dalam hal implementasi, ISC menunjukkan kinerja yang lebih baik saat dibandingkan dengan metode pengukuran kesamaan lainnya pada data berdimensi tinggi (Sohangir & Wang, 2017).

$$ISC(x, y) = \frac{\sum_{i=1}^m \sqrt{x_i y_i}}{\sqrt{(\sum_{i=1}^m x_i)} \sqrt{(\sum_{i=1}^m y_i)}} \quad (2.7)$$

Adapun keterangan dari Persamaan 2.7:

- $x$  = Vektor yang mewakili dokumen pertama
- $y$  = Vektor yang mewakili dokumen kedua
- $x_i$  = Bobot pada *term* ke- $i$  pada vektor  $x$
- $y_i$  = Bobot pada *term* ke- $i$  pada vektor  $y$
- $i$  = Indeks *term* dalam suatu kalimat
- $m$  = Jumlah total *term* dalam vektor vektor  $x$  dan  $y$

### **2.2.7 Cosine Similarity**

Tujuan dari *Cosine Similarity* adalah mendapatkan nilai similaritas dari setiap dokumen yang dibandingkan dengan mengukur kosinus sudut antara dua vektor, dengan fokus pada arah vektor daripada besarnya (Jawale, et al., 2024). Dalam kemiripan teks, setiap vektor mewakili sebuah dokumen, dan elemen-elemennya adalah frekuensi kata. (Jawale, et al., 2024).

$$\text{CosSim}(x, y) = \frac{\sum_{i=1}^m x_i \cdot y_i}{\sqrt{\sum_{i=1}^m x_i^2} \cdot \sqrt{\sum_{i=1}^m y_i^2}} \quad (2.8)$$

Keterangan dari Persamaan 2.8:

- $x$  = Vektor yang mewakili dokumen pertama
- $y$  = Vektor yang mewakili dokumen kedua
- $x_i$  = Bobot pada *term* ke- $i$  pada vektor  $x$
- $y_i$  = Bobot pada *term* ke- $i$  pada vektor  $y$
- $i$  = Indeks *term* dalam suatu kalimat
- $m$  = Jumlah total *term* dalam vektor vektor  $x$  dan  $y$

### **2.2.8 Human-Level Performance**

Meskipun algoritma dapat menunjukkan kinerja yang sangat baik, perlu dipastikan bahwa perbandingan antara kinerja manusia dan algoritma dilakukan secara adil dan tepat agar hasil yang diperoleh dapat dipercaya (Cowley, et al., 2022). Memahami bagaimana manusia menyelesaikan tugas tertentu dapat memberikan informasi yang berguna bagi penelitian di bidang *machine learning* dan *artificial intelligence* (Cowley, et al., 2022). Penting untuk mempertimbangkan apakah suatu sistem harus mencapai kinerja setara dengan manusia untuk dianggap cerdas dan apakah mesin harus menyelesaikan masalah dengan cara yang mirip dengan manusia, sehingga dapat menunjukkan pola keberhasilan dan kesalahan yang serupa (Cowley, et al., 2022). Dengan demikian, mengetahui cara manusia menyelesaikan tugas dapat membantu dalam pengembangan algoritma yang lebih baik dan lebih efektif (Cowley, et al., 2022).

### **2.2.9 Spearman Rank Correlation Coefficient (SRCC)**

*Spearman Rank Correlation Coefficient* (SRCC) adalah versi nonparametrik dari koefisien *Pearson Correlation* yang digunakan untuk menyelidiki hubungan linear antara dua variabel, khususnya pada data ordinal (Temizhan, et al., 2022). SRCC cocok digunakan ketika data tidak memenuhi asumsi parametrik, ukuran sampel kecil, atau terdapat masalah *outlier* (Temizhan, et al., 2022). Koefisien ini dapat diinterpretasikan dalam hal variabilitas peringkat dan dapat menilai hubungan monoton, di mana satu variabel cenderung naik atau turun seiring perubahan variabel lainnya (Temizhan, et al., 2022). Meskipun nonparametrik, asumsi penting untuk menggunakan SRCC adalah data harus setidaknya bersifat ordinal dan harus ada hubungan monoton antara skor pada satu variabel dengan variabel lainnya (Temizhan, et al., 2022).

$$SRCC = 1 - \frac{6 \sum d_i}{n(n^2-1)} \quad (2.9)$$

Adapun keterangan dari Persamaan 2.9:

$d_i$  = Selisih antara peringkat variabel, dihitung sebagai  $X_i - Y_i$

$n$  = Jumlah total pasangan data yang digunakan dalam perhitungan

Rentang nilai koefisien korelasi berkisar dari -1 hingga 1 (Hermanto & Harliana, 2024). Nilai 1 menunjukkan korelasi positif sempurna, nilai -1 menunjukkan korelasi negatif sempurna, dan nilai 0 menandakan tidak ada korelasi (Hermanto & Harliana, 2024). Interpretasi koefisien korelasi pada buku oleh Robert Kurniawan (2016), koefisien korelasi berkisar antara 0,00 hingga 1,00 (Hermanto & Harliana, 2024). Nilai 0,00 hingga 0,19 menunjukkan korelasi sangat lemah; 0,20 hingga 0,39 menunjukkan korelasi lemah; 0,40 hingga 0,59 menunjukkan korelasi sedang; 0,60 hingga 0,79 menunjukkan korelasi kuat; dan 0,80 hingga 1,00 menunjukkan korelasi sangat kuat (Hermanto & Harliana, 2024).

## **BAB 3 METODOLOGI**

Terdapat beberapa bagian di bab metodologi penelitian, seperti tipe penelitian, strategi penelitian, lokasi penelitian, metode pengumpulan data, metode analisis data, peralatan pendukung, dan perancangan algoritma.

### **3.1 Tipe Penelitian**

Penelitian ini merupakan penelitian non implementatif-analitik yang berarti produk yang dihasilkan berupa hasil analisis yang relevan dengan topik yang diteliti. Dalam penelitian ini, metode yang digunakan untuk menentukan resume yang paling sesuai dengan kualifikasi perekrut adalah metode perhitungan nilai similaritas tertinggi antara kualifikasi yang tercantum dalam resume dan kualifikasi yang dibutuhkan oleh perekrut sehingga dapat membantu perekrut dalam memilih 5 resume dengan tingkat kesesuaian tertinggi.

### **3.2 Strategi Penelitian**

Strategi penelitian yang dilakukan pada penelitian ini adalah penelitian eksperimen. Penelitian eksperimen adalah salah satu metode penelitian yang dapat menguji hipotesis mengenai hubungan sebab-akibat (Guritno, et al., 2011, p. 29). Kemudian, didefinisikan juga bahwa pendekatan ini merupakan penelitian untuk menguji sebab akibat antar variabel melalui langkah manipulasi, pengendalian, dan pengamatan (Musfiqon, 2016, p. 60). Penelitian eksperimen dilaksanakan dengan maksud mengetahui akibat dari suatu perlakuan melalui cara sengaja menimbulkan kejadian (eksperimen) (Effendi, 2013, p. 88).

### **3.3 Lokasi Penelitian**

Penelitian ini akan dilaksanakan di Fakultas Ilmu Komputer, Universitas Brawijaya, Kota Malang, Jawa Timur.

### **3.4 Metode Pengumpulan Data**

Data yang digunakan dalam penelitian ini didapatkan dari *platform* Kaggle, berjudul "Resume Dataset" yang dibuat oleh Snehaan Bhawal. Dataset ini terdiri dari 2.484 resume yang dikategorikan berdasarkan jenis pekerjaan yang dilamar, seperti HR, Desainer, Teknologi Informasi, Guru, dan kategori lainnya. *Dataset* tersebut mencakup format resume dalam bentuk *string* (teks) dan setiap resume diidentifikasi dengan ID unik. Informasi yang terdapat dalam dataset meliputi teks resume, data HTML hasil *web scraping*, dan kategori pekerjaan (Bhawal, 2021).

### **3.5 Metode Analisis Data**

Tujuan menganalisis data, antara lain mendapatkan perasaan terhadap data, menguji kualitas data, dan menguji hipotesis penelitian (Guritno, et al., 2011, p. 183). Menurut Cholissodin & Riyandani (2016), terdapat beberapa fase pada gambaran umum siklus hidup analitik data, seperti:

1. *Discovery*. Fase ini meliputi proses belajar, mencari dan menyelidiki fakta-fakta, mengidentifikasi masalah, mengembangkan konteks dan pemahaman, dan belajar tentang sumber data yang dibutuhkan, diikuti dengan perumusan hipotesis awal yang nantinya dapat diuji dengan data (Cholissodin & Riyandani, 2018, p. 22).
2. *Data Preparation*. Fase ini meliputi persiapan data sebelum dipakai untuk proses *modelling* dan *evaluation* yang dibagi menjadi dua bagian, yakni *cleaning* untuk menyeleksi beberapa fitur dan *transformation* untuk mengubah bentuk data ke dalam bentuk yang bisa diterima oleh algoritma (Abdusyukur, 2023).
3. *Model Planning*. Fase ini merupakan proses penentuan metode, teknik, dan alur kerja dengan mengeksplorasi data untuk mempelajari hubungan antara variabel yang selanjutnya memilih variabel kunci dan model yang paling cocok untuk digunakan (Cholissodin & Riyandani, 2018, p. 23).
4. *Model Building*. Pada fase ini, *dataset* dikembangkan untuk pengujian, pelatihan, dan tujuan produksi, serta mempertimbangkan apakah dengan alat yang ada akan cukup untuk menjalankan model (Cholissodin & Riyandani, 2018, p. 23).
5. *Communicate Result*. Pada fase ini, temuan-temuan yang didapatkan akan didiskusikan dengan para pemangku kepentingan untuk menentukan apakah hasil proyek tersebut sukses atau mengalami kegagalan (Cholissodin & Riyandani, 2018, p. 24).
6. *Operationalize*. Fase ini merupakan yang terakhir dengan menyerahkan laporan akhir, pengarahan, kode, dan dokumen teknis (Cholissodin & Riyandani, 2018, p. 24).

### **3.6 Metode Evaluasi**

Evaluasi metode similaritas teks dilakukan dengan menggunakan *human-level performance* sebagai tolak ukur untuk membandingkan korelasi antara keluaran lima resume dengan nilai similaritas tertinggi yang dihasilkan oleh implementasi metode dengan peringkat *ground truth* keluaran lima resume tersebut oleh seorang ahli yang memiliki pengalaman rekrutmen selama 2 tahun dan telah meninjau lebih dari 5000 resume di bidang *sales*, *marketing*, teknologi, *healthcare*, *accounting*, *finance*, *human resources*, dan *legal*. Evaluasi ini dilakukan untuk setiap posisi lowongan kerja dari total 24 kualifikasi lowongan kerja dan hasilnya dianalisis menggunakan tiga parameter penilaian, yakni korelasi sebagai parameter utama, serta relevansi dan senioritas sebagai parameter tambahan.

### **3.7 Peralatan Pendukung**

Dalam melakukan penelitian ini dari awal hingga akhir, diperlukan beberapa peralatan pendukung untuk membantu kelancaran jalannya penelitian. Peralatan pendukung tersebut meliputi perangkat lunak (*software*) dan perangkat keras (*hardware*)

#### **3.7.1 Perangkat Lunak (*Software*)**

Perangkat lunak yang digunakan, antara lain:

1. Sistem operasi Microsoft Windows 10 Home 64-bit
2. Jupyter Notebook Versi 7.0.8
3. Bahasa pemrograman Python 3.12.4
4. *Library* Python Pandas Versi 2.2.3
5. *Library* Python BeautifulSoup4 (bs4) Versi 4.12.3
6. *Library* Python Gensim Versi 4.3.3
7. *Library* Python Numpy Versi 1.26.4
8. *Library* Python Scikit-learn Versi 1.6.1
9. *Library* Python NLTK Versi 3.9.1
10. *Library* Python TQDM Versi 4.67.1
11. Microsoft® Word 2016 MSO (*Version* 2505 *Build* 16.0.18827.20102)  
32-bit

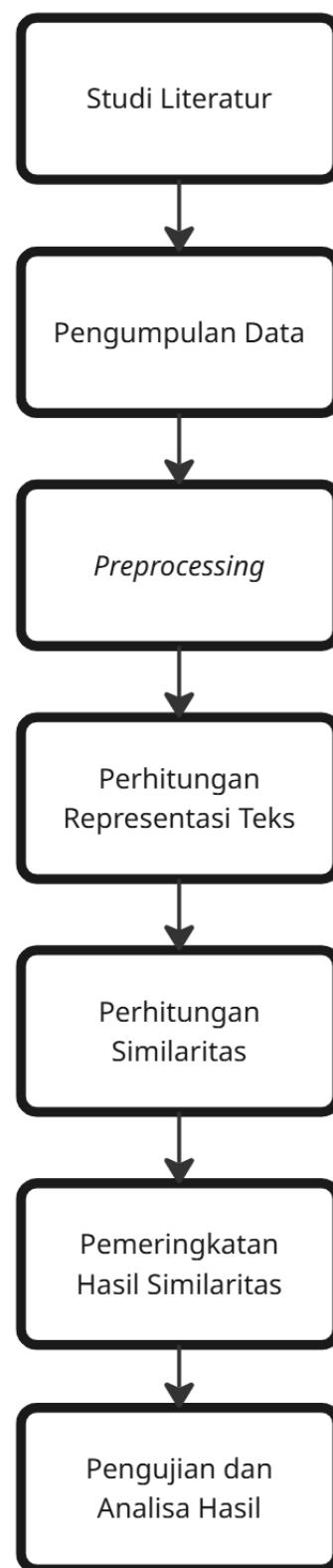
#### **3.7.2 Perangkat Keras (*Hardware*)**

Perangkat keras yang digunakan, antara lain:

1. Windows 10 Home (2009)
2. Intel(R) Core(TM) i5-6200U CPU @ 2.30GHz, 2401 Mhz, 2 Core(s), 4 Logical Processor(s)
3. Memori RAM 8,00 GB
4. SSD SanDisk Z400s 2.5 7MM 256GB

### **3.8 Perancangan Algoritma**

Pada perancangan algoritma dijabarkan proses pengimplementasian metode similaritas teks untuk otomatisasi penyaringan resume berdasarkan kualifikasi yang instansi butuhkan. Sebelumnya, dilakukan studi literatur, lalu proses implementasi ini dimulai dengan melakukan pra-pemrosesan *dataset* resume dan kualifikasi lowongan kerja. Kemudian, dilakukan perhitungan representasi teks untuk keduanya yang menghasilkan vektor-vektor guna menghitung similaritas antara vektor resume dengan vektor dari kualifikasi lowongan kerja. Setelah skor similaritas didapatkan, resume pun diurutkan mulai dari yang paling besar berdasarkan skor similaritasnya dan diberikan peringkat agar dapat dilakukan pengujian.



**Gambar 3.1 Struktur proses implementasi pemeringkatan similaritas resume dan kualifikasi lowongan kerja**

## BAB 4 PERANCANGAN

Bab perancangan menjelaskan tentang perancangan algoritma dari metode-metode serta *flow diagram* dari setiap algoritma yang digunakan pada penelitian ini. Selain itu, bab ini juga merincikan perhitungan manual.

### 4.1 Deskripsi Umum

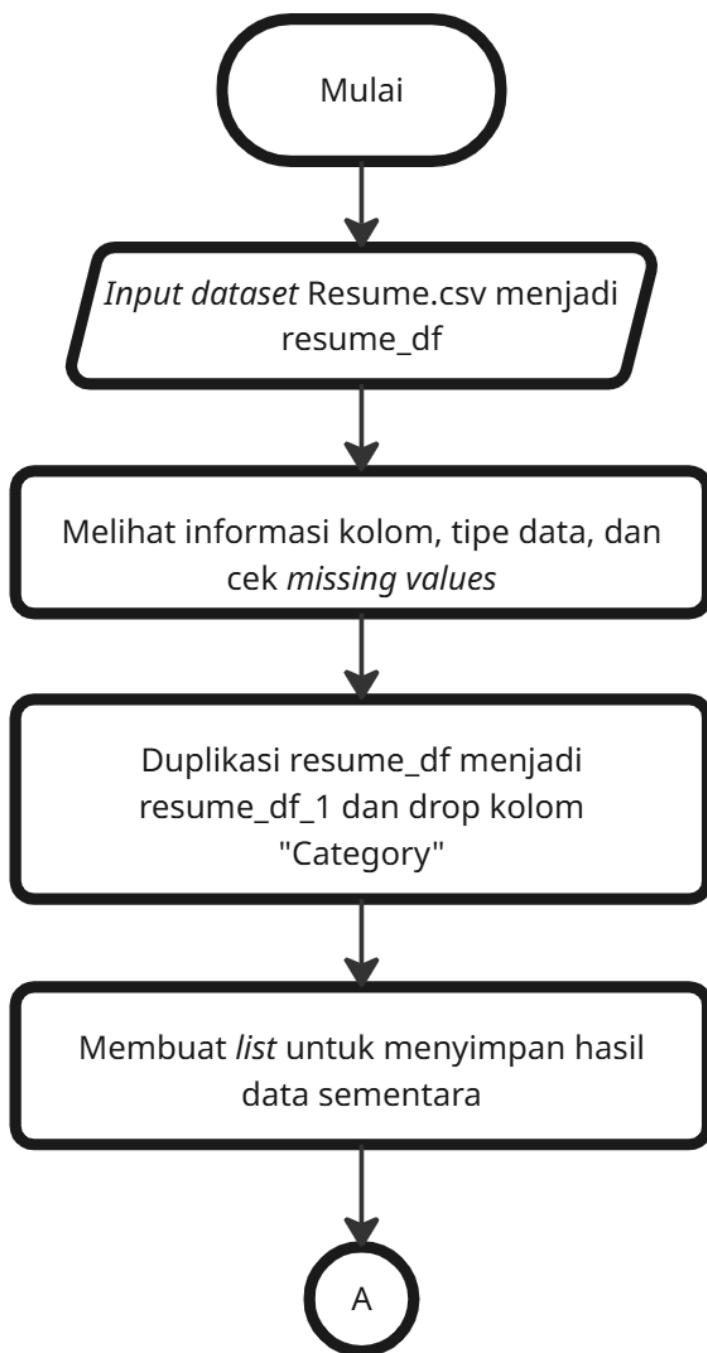
Penelitian ini dilakukan dengan memeringkat lima resume dari *dataset* berdasarkan skor similaritas dengan kualifikasi suatu lowongan kerja dalam format CSV. Langkah pertama yang dilakukan adalah ekstraksi bagian-bagian (*section*) dari resume, diikuti dengan *preprocessing* pada *dataset* resume dan kualifikasi lowongan kerja. Setelah data dibersihkan, dilakukan perhitungan representasi teks untuk *dataset* resume dan kualifikasi lowongan kerja. Kemudian, vektor yang diperoleh dari resume dan kualifikasi lowongan kerja dilakukan perhitungan similaritas. Pemeringkatan lima resume ditentukan dengan skor similaritas tertinggi untuk setiap kualifikasi lowongan kerja. Hasil pemeringkatan diberikan ke seorang ahli untuk dievaluasi secara *human-level performance*.

Pada proses pemeringkatan lima resume, terdapat dua masukan, yakni *dataset* resume (Resume.csv) dan kualifikasi lowongan kerja (kualifikasi\_loker.csv) yang merupakan kumpulan 24 kualifikasi lowongan kerja dari situs pencarian kerja <https://id.jobstreet.com/>. Dalam *dataset* resume, terdapat kolom ID, Resume\_str, Resume\_html, dan Category. Kolom ID merupakan nomor *identifier* yang dimiliki setiap resume, kolom Resume\_str merupakan isi dari resume, kolom Resume\_html merupakan isi dari resume dengan format HTML, dan kolom Category merupakan pengelompokan industri atau bidang posisi pekerjaan dari setiap resume. Diketahui terdapat 24 kategori industri, yaitu: "HR", "DESIGNER", "INFORMATION-TECHNOLOGY", "TEACHER", "ADVOCATE", "BUSINESS-DEVELOPMENT", "HEALTHCARE", "FITNESS", "AGRICULTURE", "BPO", "SALES", "CONSULTANT", "DIGITAL-MEDIA", "AUTOMOBILE", "CHEF", "FINANCE", "APPAREL", "ENGINEERING", "ACCOUNTANT", "CONSTRUCTION", "PUBLIC-RELATIONS", "BANKING, ARTS", dan "AVIATION". Kategori-kategori tersebut yang dijadikan landasan untuk mencari 24 kualifikasi lowongan kerja.

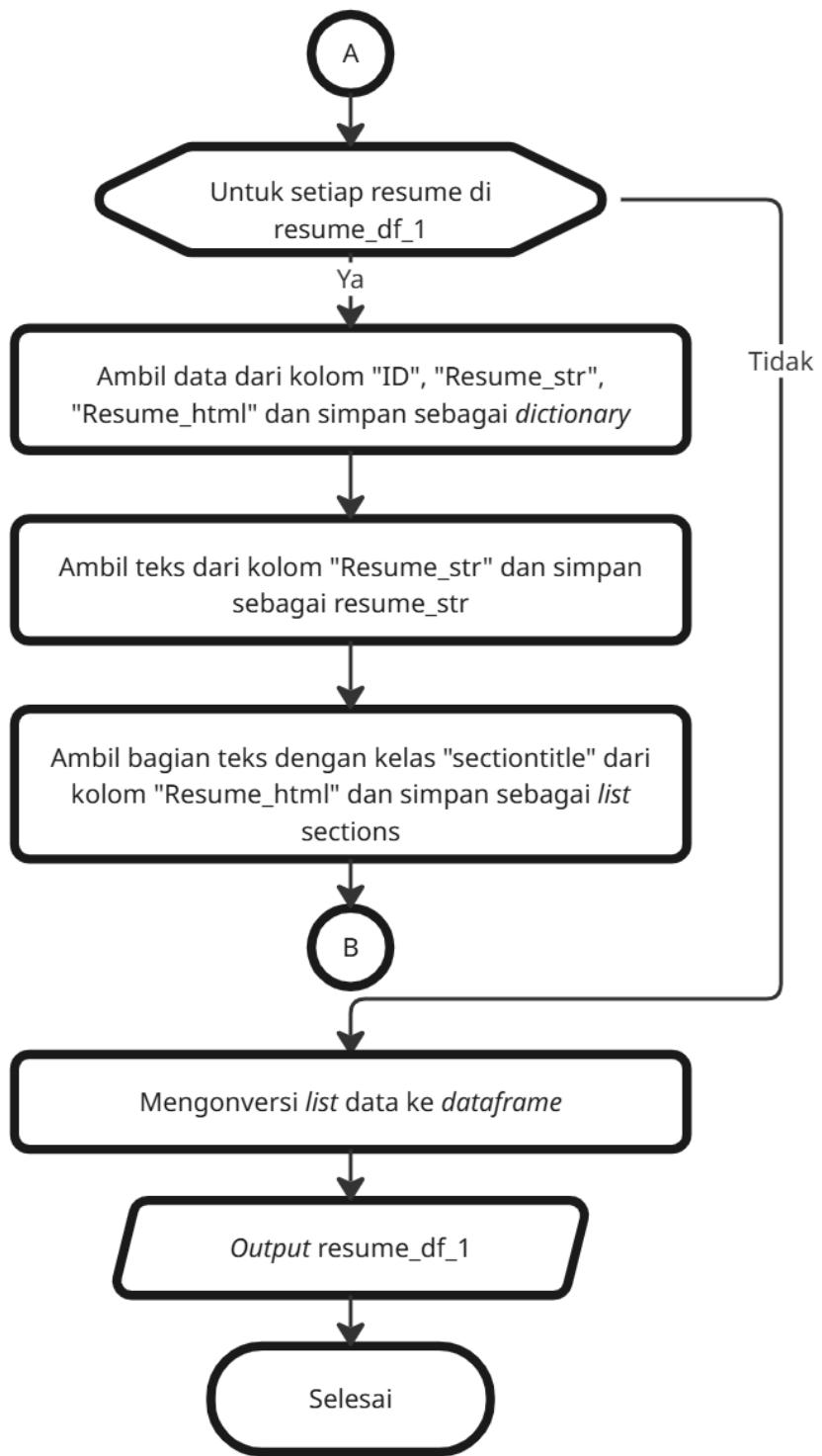
### 4.2 *Preprocessing*

#### 4.2.1 Ekstraksi Section

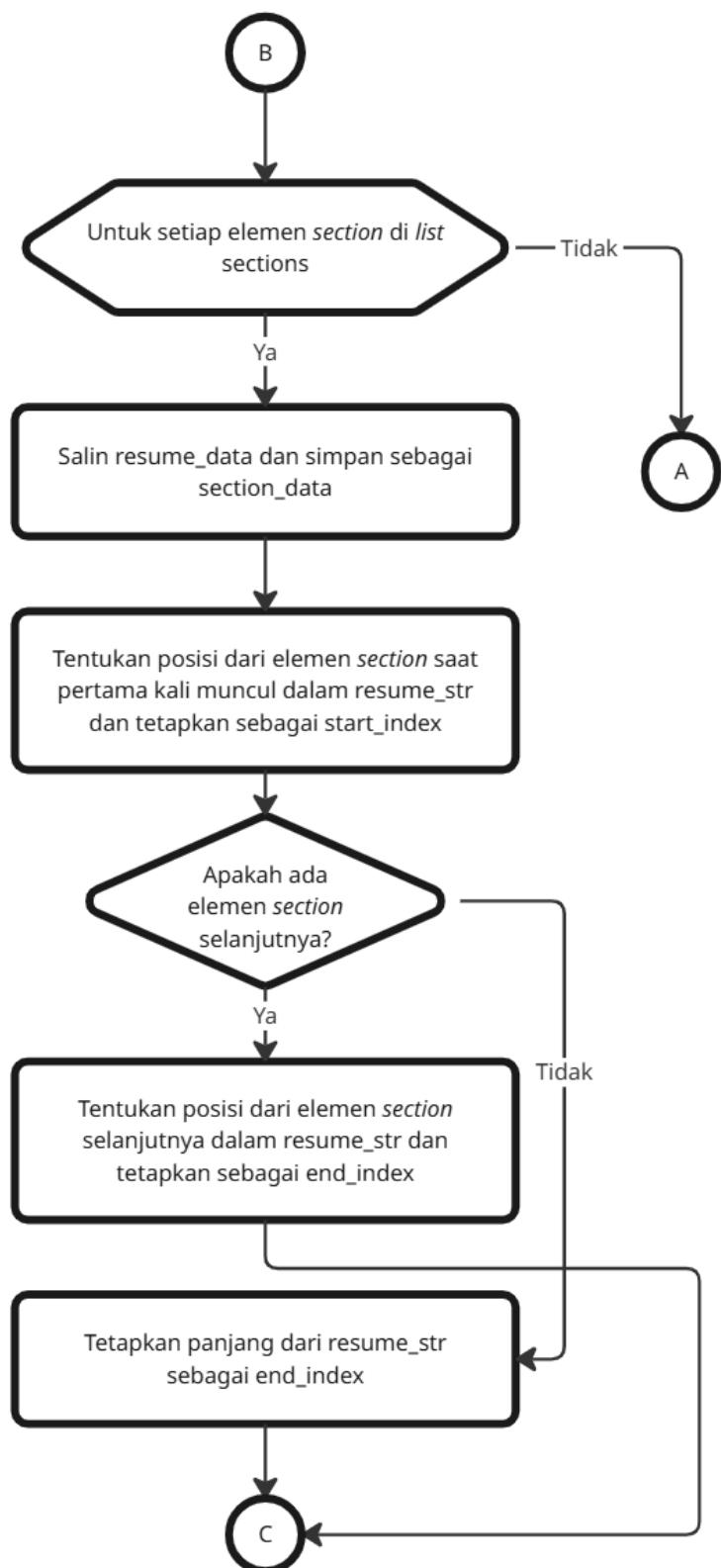
Langkah pertama adalah men-*drop* kolom yang tidak digunakan seperti kolom Category dan dilanjut melakukan ekstraksi bagian-bagian (*section*) dengan mengambil *class sectiontitle* dari kolom Resume\_html. Setelah *section* setiap resume diketahui, isian dari setiap bagian diambil dari kolom Resume\_str. Hasilnya adalah DataFrame resume\_df\_1 yang memiliki kolom ID, Resume\_str, Resume\_hml, Section, dan Text. Diagram alur proses ekstraksi *section* tertera pada Gambar 4.1 hingga Gambar 4.4.



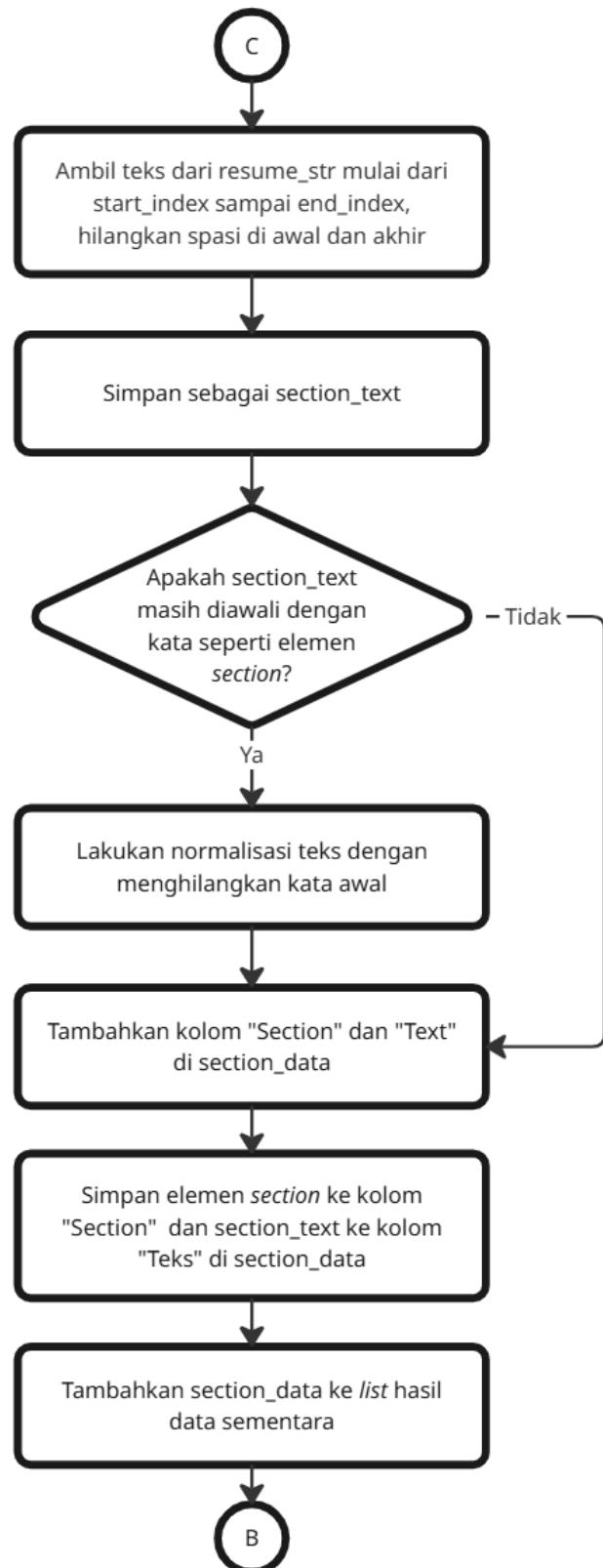
Gambar 4.1 Diagram alur ekstraksi *section*



Gambar 4.2 Diagram alur ekstraksi *section*



Gambar 4.3 Diagram alur ekstraksi *section*



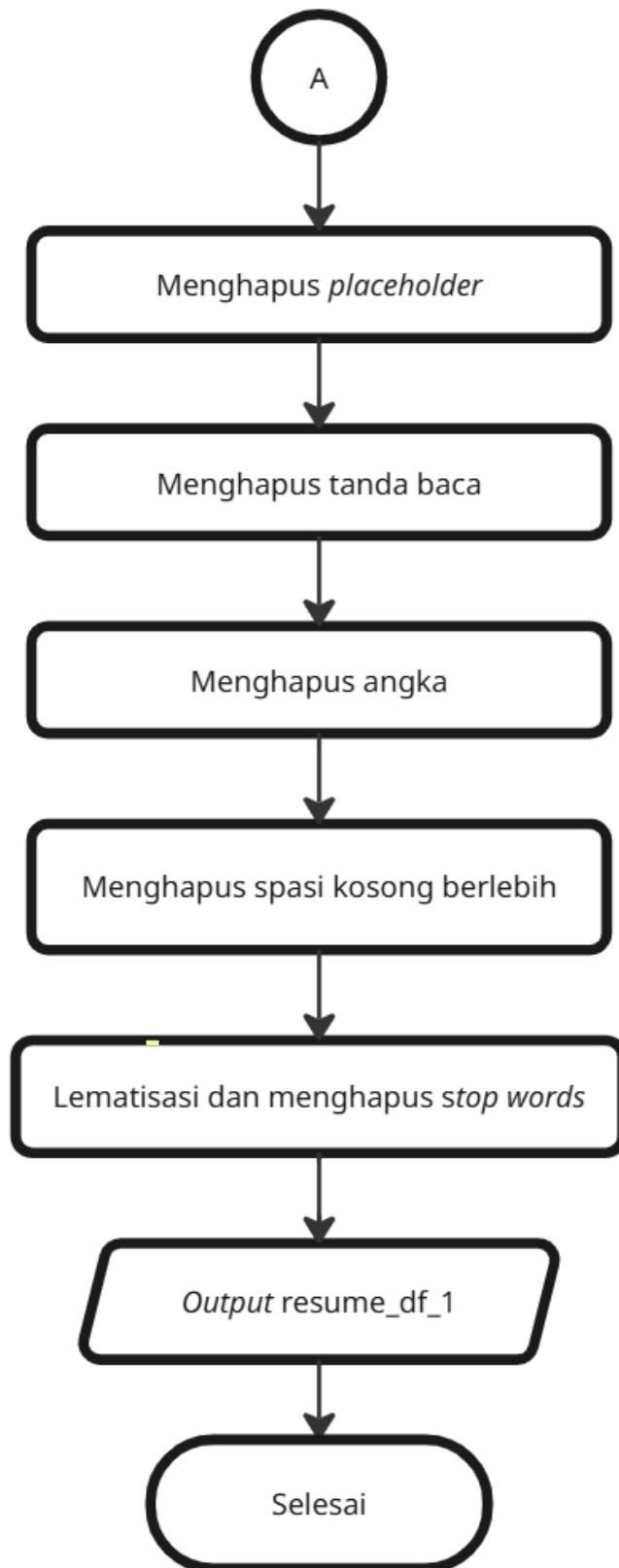
**Gambar 4.4 Diagram alur ekstraksi *section***

#### 4.2.2 Preprocessing Isian ResUME

Setelah langkah ekstraksi *section* dilakukan dan menghasilkan DataFrame `resume_df_1`, selanjutnya adalah langkah *preprocessing* untuk kolom Text yang merupakan isian dari setiap *section* pada setiap resume. Langkah-langkah yang dilakukan mencakup penghapusan *email*, nomor telepon, tanggal, dan tahun. Selain itu, juga dilakukan penghapusan kata “*Current*”, “*Present*”, penghapusan istilah-istilah *placeholder*, penghapusan tanda baca, penghapusan tanggal, penghapusan angka, dan penghapusan spasi berlebih. Selanjutnya, dilakukan lematisasi dan penghapusan *stopword* agar kata-kata ditransformasikan menjadi bentuk dasar sesuai dengan kamus. Diagram alur proses *preprocessing* isian resume tertera pada Gambar 4.5 hingga Gambar 4.6.



Gambar 4.5 Diagram alur *preprocessing* isian resume



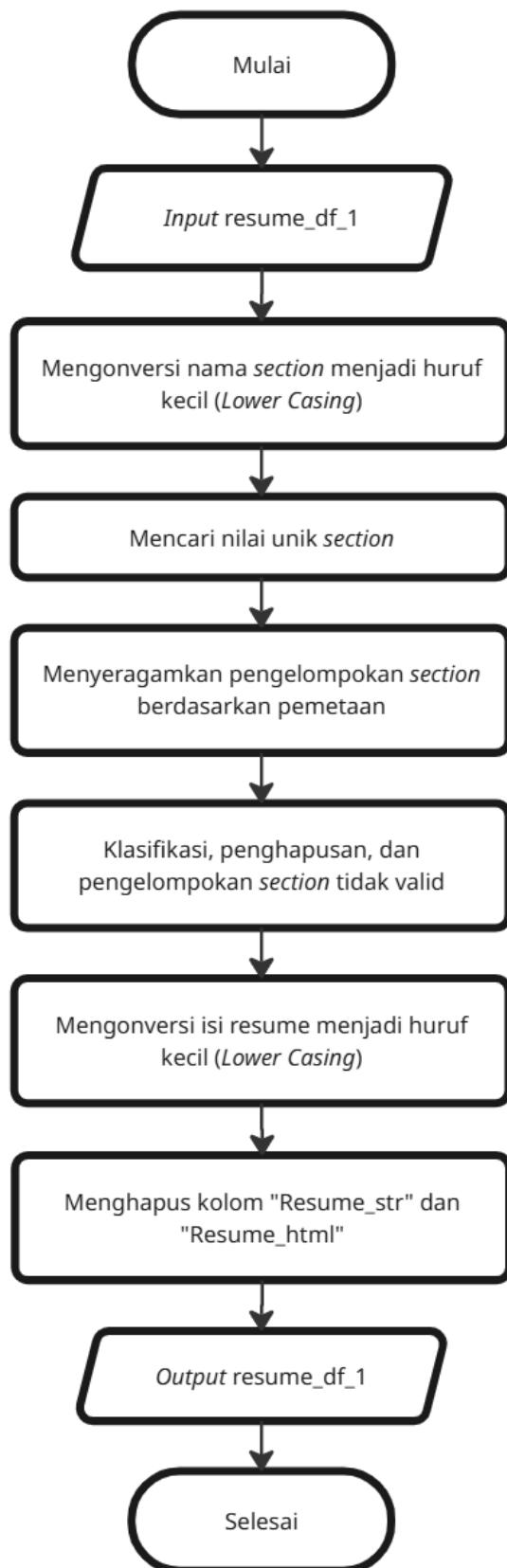
Gambar 4.6 Diagram alur *preprocessing* isian resume

#### **4.2.3 Preprocessing Penamaan Section**

Setiap resume umumnya memiliki struktur informasi yang serupa, seperti bagian *Education*, *Work Experience*, *Skills*, dan *Summary*. Namun, penamaan atau label dari setiap bagian tersebut dapat sangat bervariasi antar resume, misalnya *Work Experience* bisa juga ditulis sebagai *Experience*. Oleh karena itu, diperlukan tahap *preprocessing* penamaan *section* untuk menyamakan atau menyeragamkan nama-nama *section* tersebut ke dalam satu format baku. Berikut ini merupakan daftar *section* standar yang digunakan.

1. *Summary*. Bagian yang berisi ringkasan profil, tujuan karir, atau deskripsi singkat.
2. *Accomplishments/Awards*. Bagian yang berisi pencapaian, penghargaan, atau prestasi yang pernah diraih.
3. *Skills/Qualifications*. Bagian yang berisi daftar keterampilan teknis maupun non-teknis, serta kualifikasi lainnya.
4. *Education*. Bagian yang berisi latar belakang pendidikan formal.
5. *Experience*. Bagian yang berisi pengalaman kerja atau pengalaman profesional lainnya.
6. *Organization*. Bagian yang berisi pengalaman dalam organisasi, kepanitiaan, atau kegiatan sosial.
7. *Projects*. Bagian yang berisi proyek-proyek yang pernah dikerjakan secara individu maupun kelompok.
8. *Certifications*. Bagian yang berisi rincian sertifikasi yang diperoleh.
9. *Portfolio*. Bagian yang berisi riwayat karya atau tautan ke portofolio *online*.
10. *Others*. Bagian-bagian lain yang tidak termasuk dalam *section* di atas, seperti referensi, hobi, atau informasi tambahan lainnya.

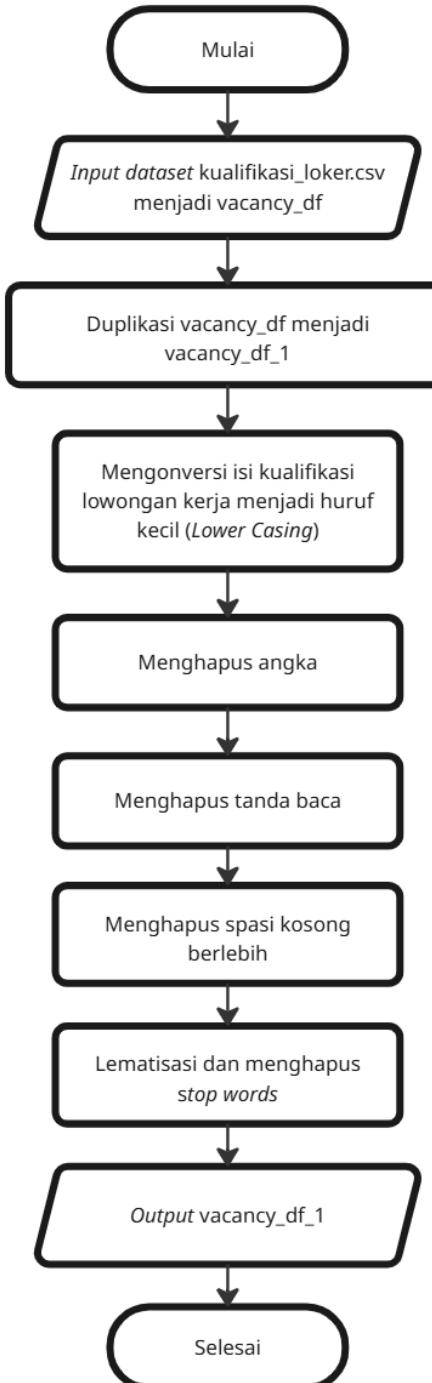
Diagram alur proses *preprocessing* penamaan *section* resume tertera pada Gambar 4.7.



Gambar 4.7 Diagram alur *preprocessing* penamaan *section* resume

#### 4.2.4 Preprocessing Kualifikasi Lowongan Kerja

Dilakukan *preprocessing* pada kolom *Description* yang berisi kebutuhan terkait suatu posisi lowongan kerja. Langkah-langkah yang dilakukan mencakup tokenisasi dan lematisasi agar kata-kata ditransformasikan menjadi bentuk dasar sesuai dengan kamus. Diagram alur proses *preprocessing* isian kualifikasi lowongan kerja tertera pada Gambar 4.8.

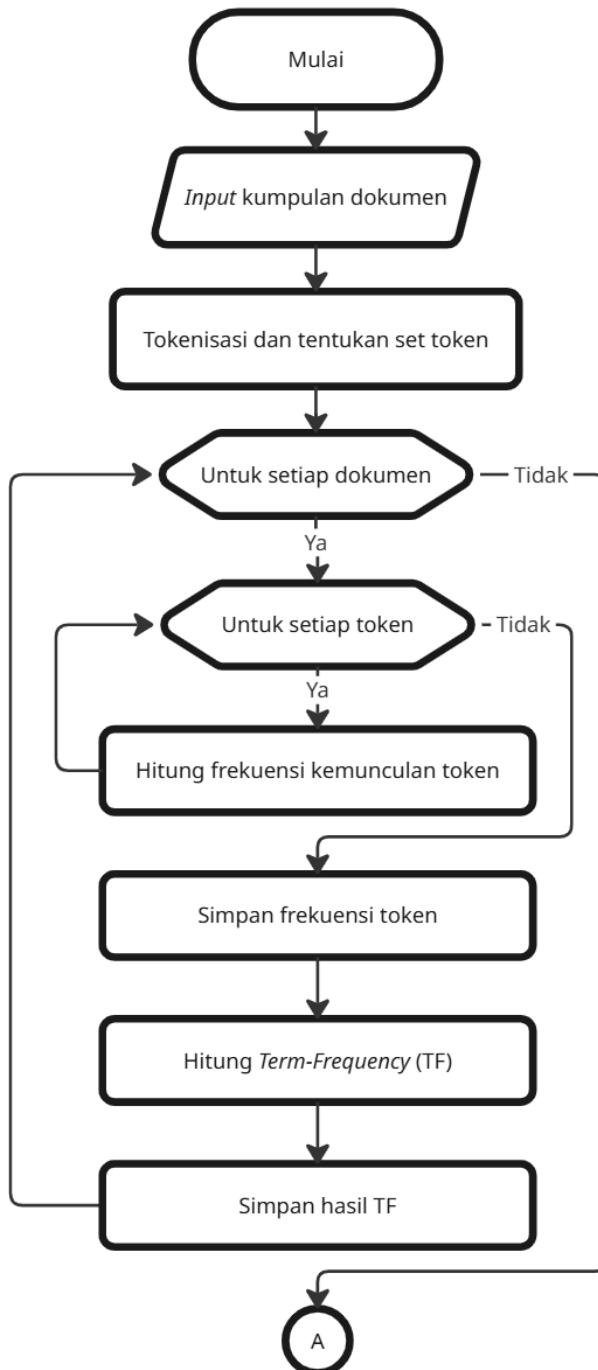


Gambar 4.8 Diagram alur *preprocessing* isian kualifikasi lowongan kerja

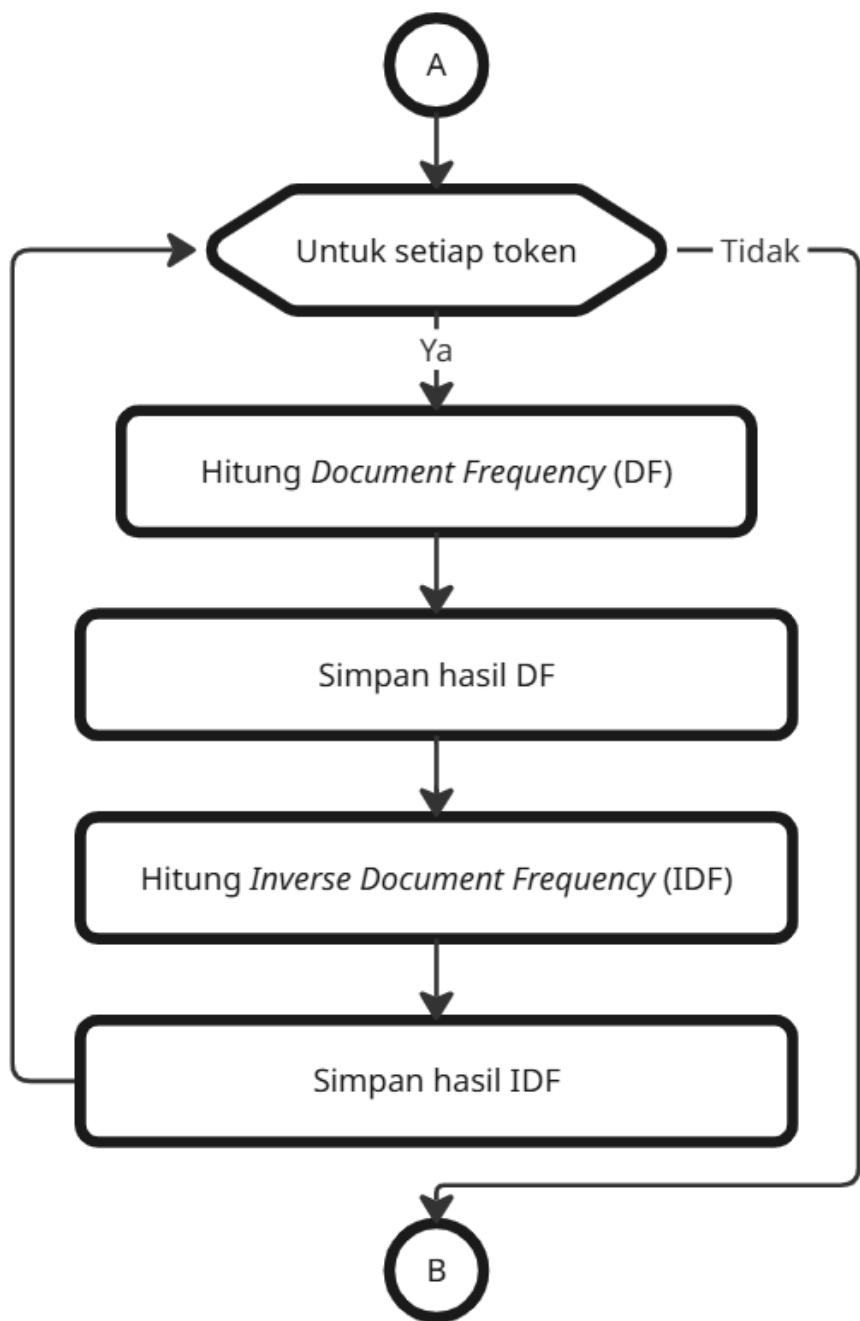
## 4.3 Perhitungan Representasi Teks

### 4.3.1 TF-IDF

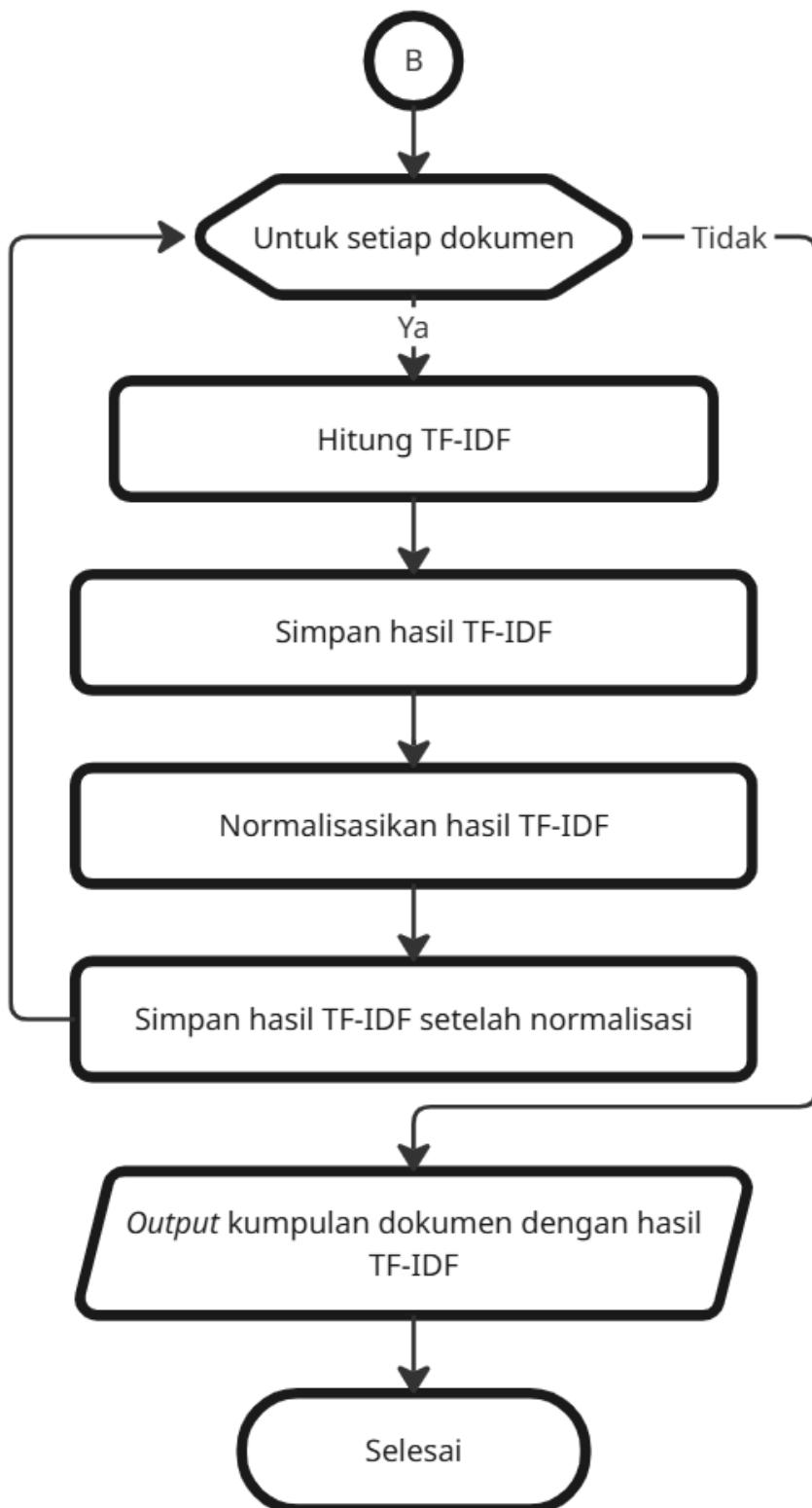
Pengimplementasian TF-IDF menggunakan *library* Scikit-learn. Meskipun proses implementasi menggunakan library, diagram alur yang menjelaskan detail perhitungan TF-IDF tertera pada Gambar 4.9 hingga Gambar 4.11.



Gambar 4.9 Diagram alur TF-IDF



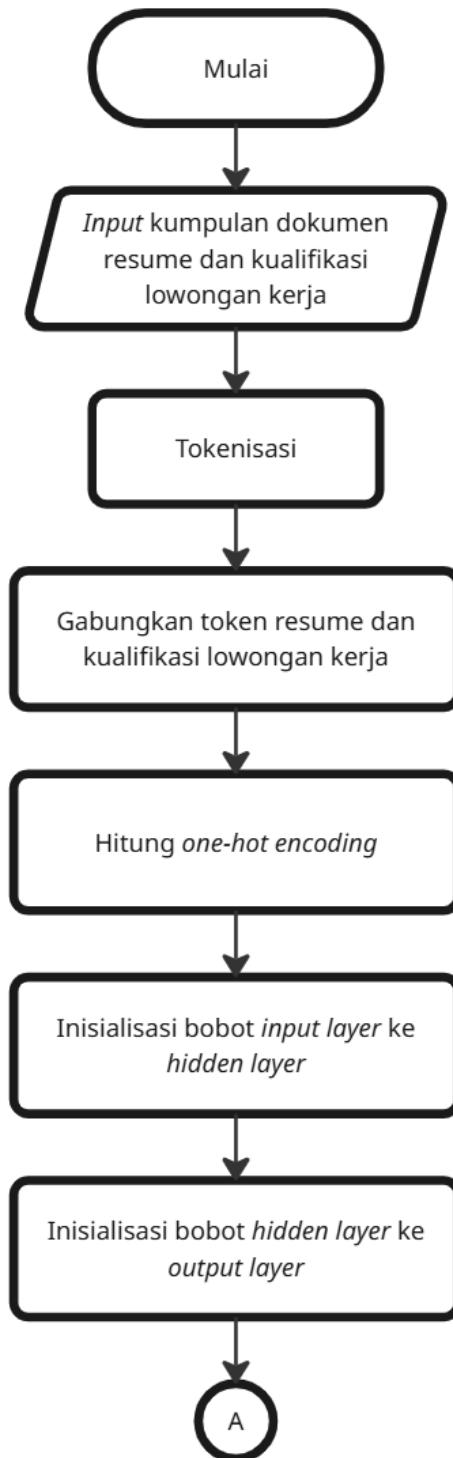
Gambar 4.10 Diagram alur TF-IDF



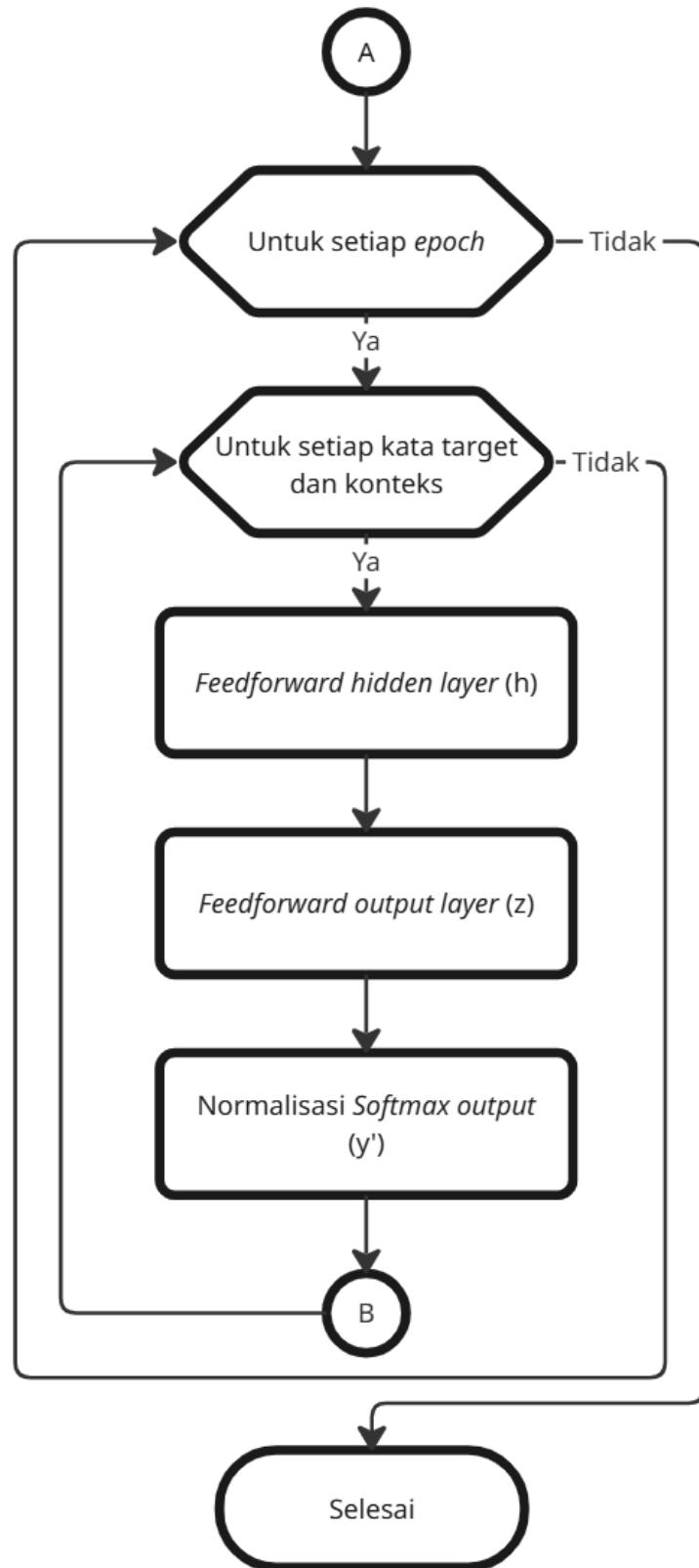
Gambar 4.11 Diagram alur TF-IDF

#### 4.3.2 Word2Vec

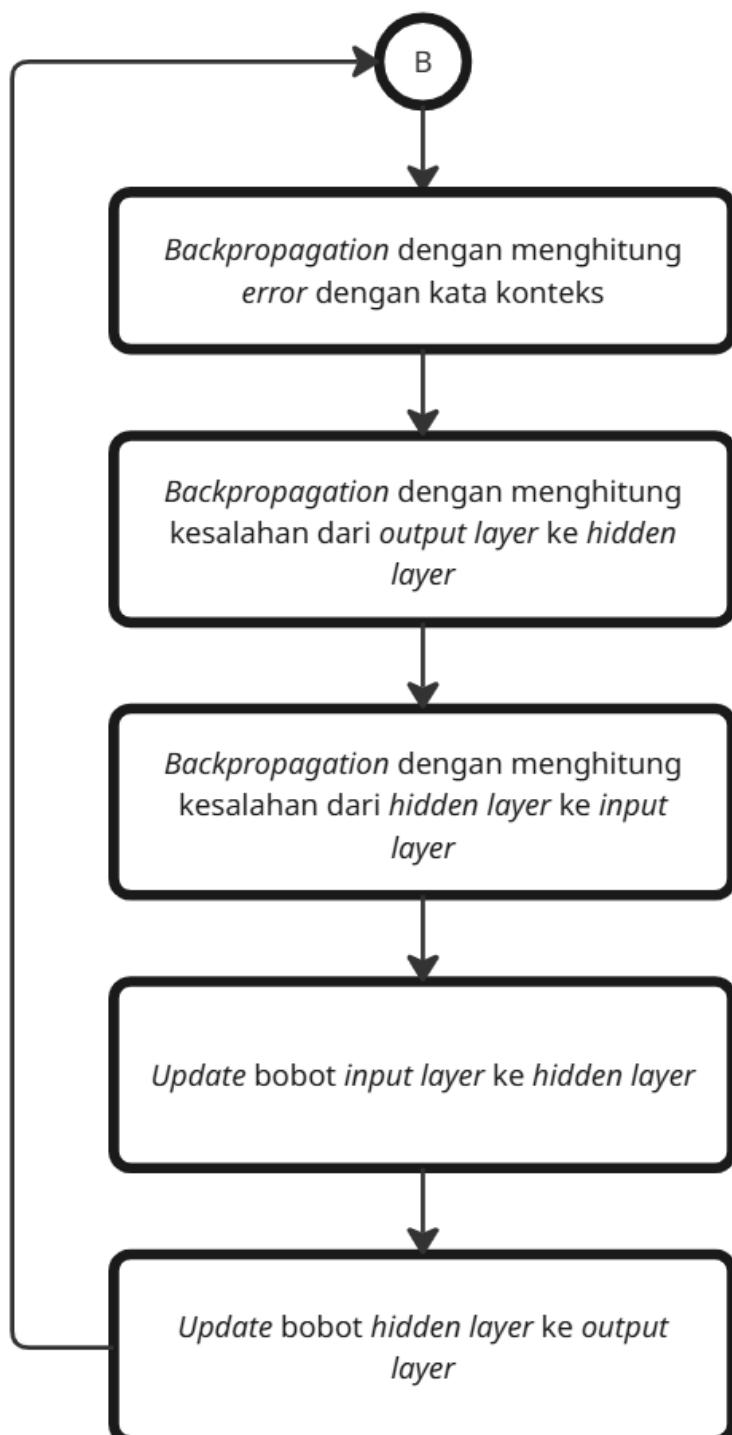
Pengimplementasian Word2Vec menggunakan *library* Gensim. Meskipun proses implementasi menggunakan *library*, diagram alur yang menjelaskan detail perhitungan Word2Vec tertera pada Gambar 4.12 hingga Gambar 4.14.



Gambar 4.12 Diagram alur Word2Vec



Gambar 4.13 Diagram alur Word2Vec

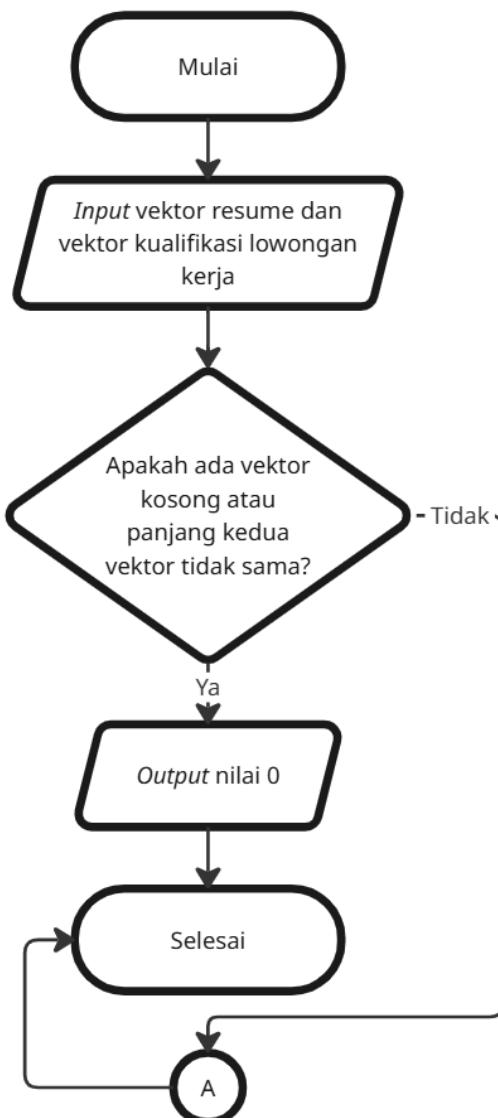


Gambar 4.14 Diagram alur Word2Vec

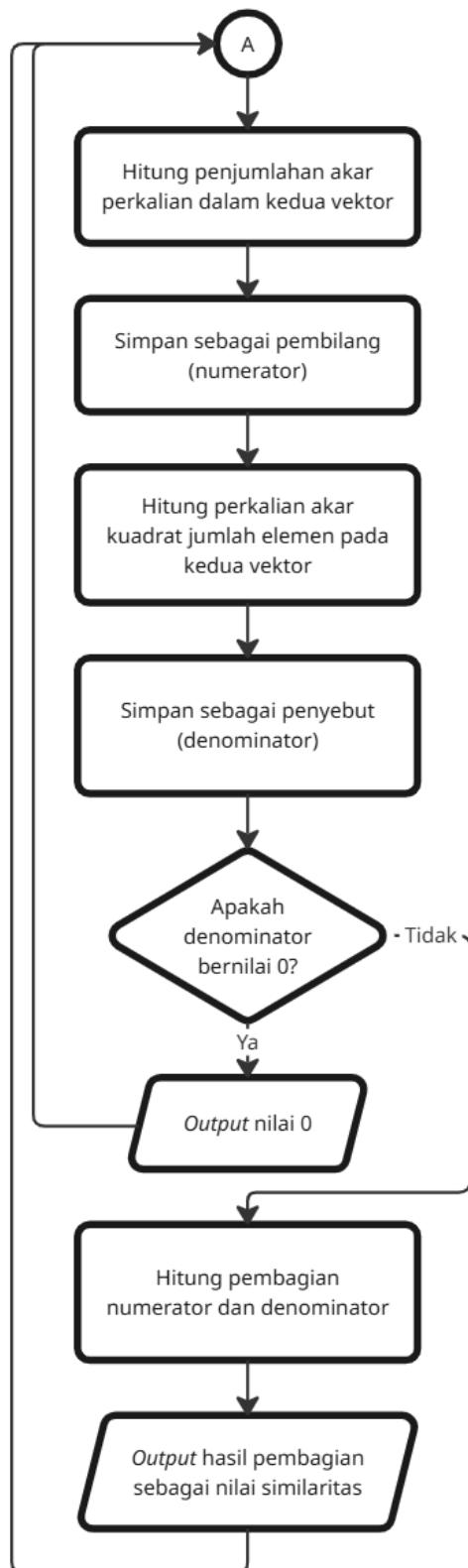
## 4.4 Perhitungan Similaritas

### 4.4.1 *Improved Sqrt-Cosine Similarity*

Proses perhitungan *Improved Sqrt-Cosine* (ISC) *Similarity* digunakan untuk mengukur seberapa mirip antara resume dengan kualifikasi lowongan kerja. Pertama, fungsi ini memeriksa apakah kedua vektor ada. Jika tidak, hasilnya adalah 0. Selanjutnya, dihitung skor similaritas dengan menjumlahkan akar dari hasil kali elemen yang bersesuaian dalam kedua vektor sebagai pembilang. Hasil tersebut dibagi dengan hasil kali akar dari total nilai di masing-masing vektor sebagai penyebut. Terakhir, fungsi ini mengembalikan hasil pembagian yang merupakan skor kemiripan kedua vektor tersebut. Diagram alur proses perhitungan similaritas ISC tertera pada Gambar 4.15 hingga 4.16.



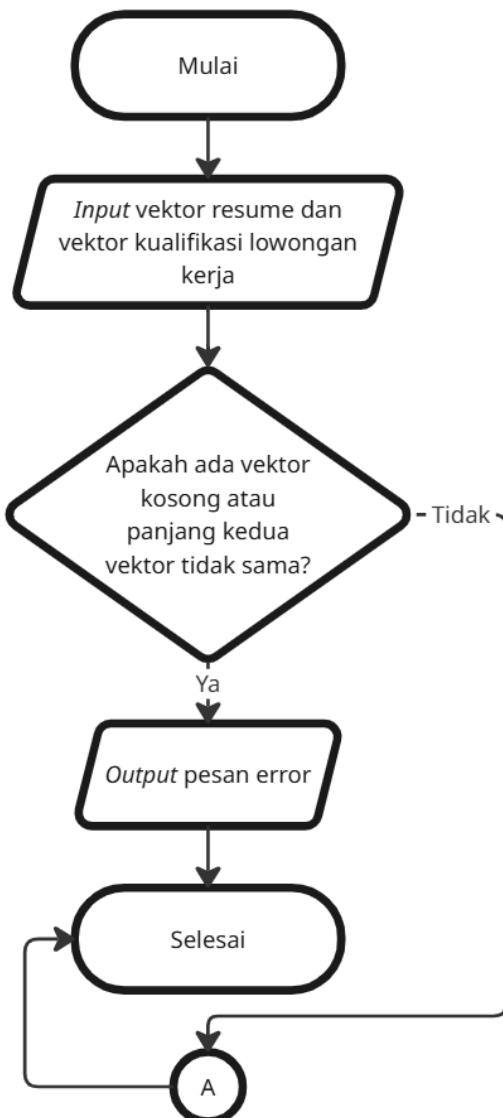
Gambar 4.15 Diagram alur *Improved Sqrt-Cosine Similarity*



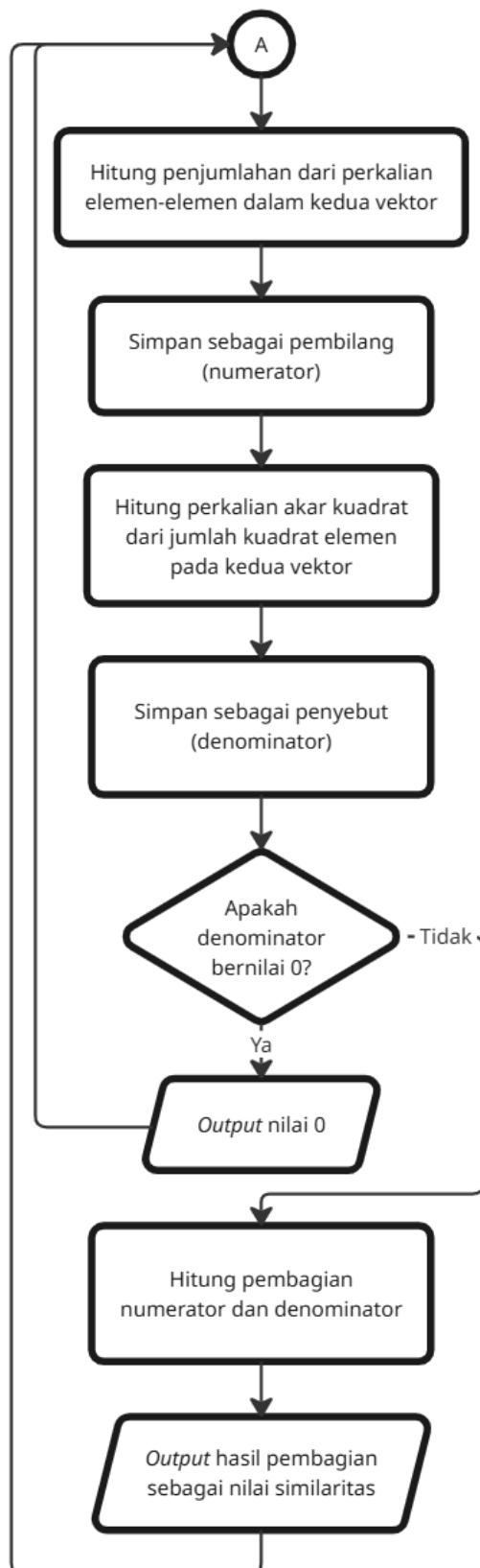
**Gambar 4.16 Diagram alur *Improved Sqrt-Cosine Similarity***

#### 4.4.2 Cosine Similarity

Meskipun dalam implementasi program digunakan *library* Scikit-learn untuk menghitung nilai *Cosine Similarity* (CosSim) antara resume dan kualifikasi lowongan kerja, proses perhitungan di balik fungsi tersebut tetap dijelaskan melalui diagram alur pada Gambar 4.17 dan 4.18. Pertama, fungsi ini akan memeriksa keberadaan kedua vektor. Jika salah satu tidak tersedia, maka akan menghasilkan pesan *error*. Selanjutnya, nilai CosSim dihitung dengan membagi jumlah hasil perkalian elemen-elemen bersesuaian dari kedua vektor (pembilang atau numerator) dengan hasil perkalian akar kuadrat dari jumlah kuadrat elemen pada masing-masing vektor (penyebut atau denominator). Skor akhir similaritas diperoleh dari pembagian antara pembilang dan penyebut. Diagram alur proses perhitungan CosSim tertera pada Gambar 4.17 hingga 4.18.



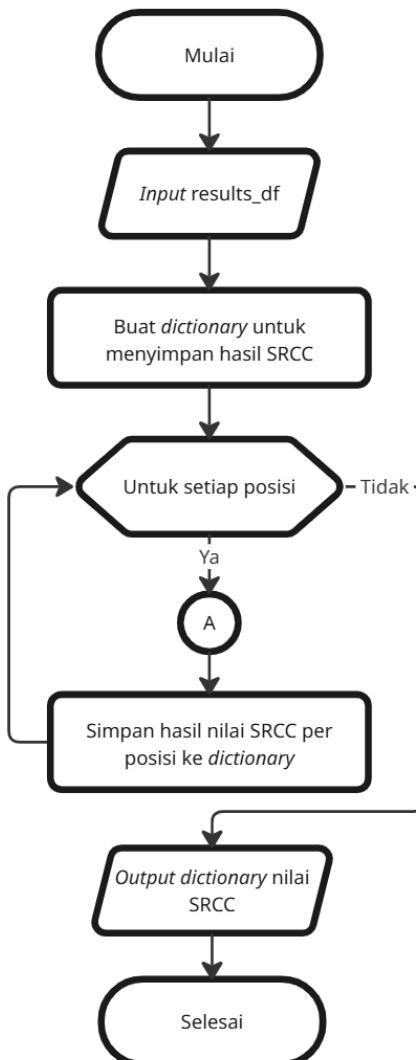
Gambar 4.17 Diagram alur *Cosine Similarity*



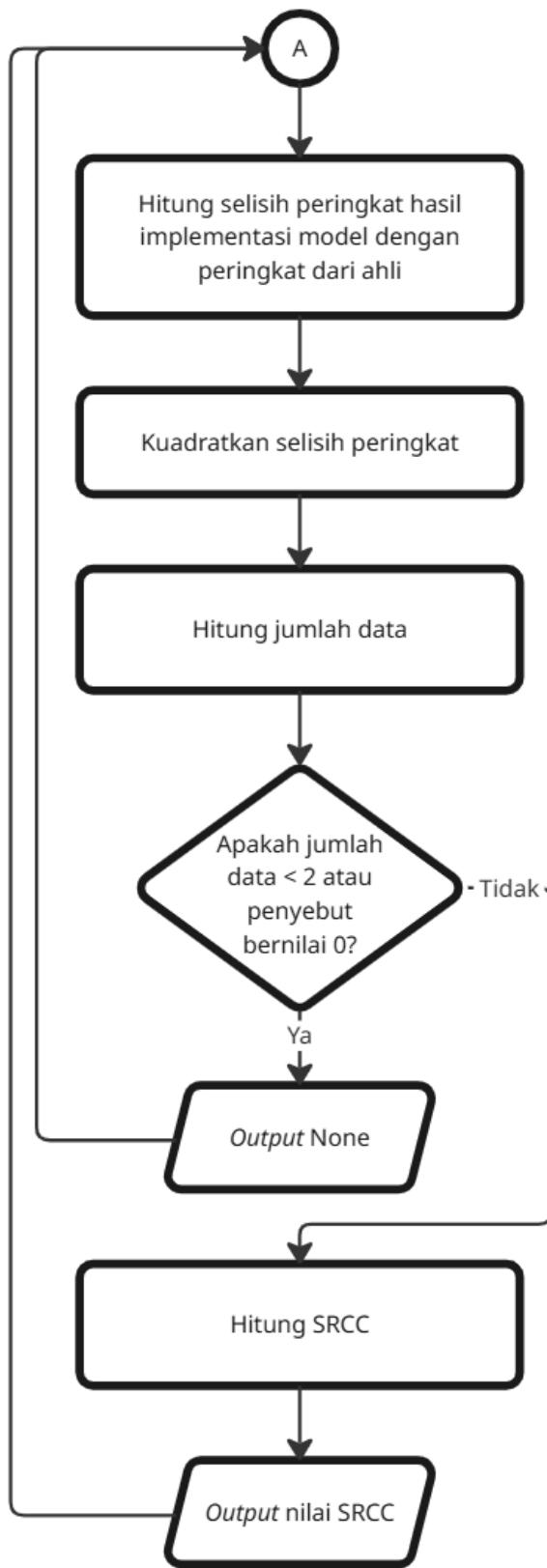
Gambar 4.18 Diagram alur *Cosine Similarity*

## 4.5 Perhitungan Korelasi

Pada proses perhitungan korelasi, *Spearman Rank Correlation Coefficient* (SRCC) digunakan untuk mengukur seberapa sesuai antara dua peringkat, yaitu peringkat hasil keluaran implementasi metode dan peringkat yang disusun oleh ahli berdasarkan evaluasi terhadap hasil keluaran implementasi metode. Pertama, DataFrame `results_df` yang berisi kolom peringkat hasil implementasi metode (`Rank`), `Resume_ID`, `Position`, `Similarity_Score`, dan peringkat *ground truth* dari ahli (`Rank_Expert`) diinput. Selanjutnya, dihitung selisih antara peringkat hasil implementasi metode dan ahli per resume untuk setiap posisi. Selisih tersebut dikuadratkan dan dijumlahkan untuk memperoleh total deviasi peringkat. Terakhir, hasil penjumlahan tersebut digunakan dalam rumus SRCC untuk mendapatkan nilai korelasi yang menunjukkan sejauh mana hasil implementasi metode sesuai dengan penilaian ahli. Diagram alur proses perhitungan SRCC tertera pada Gambar 4.19 hingga 4.20.



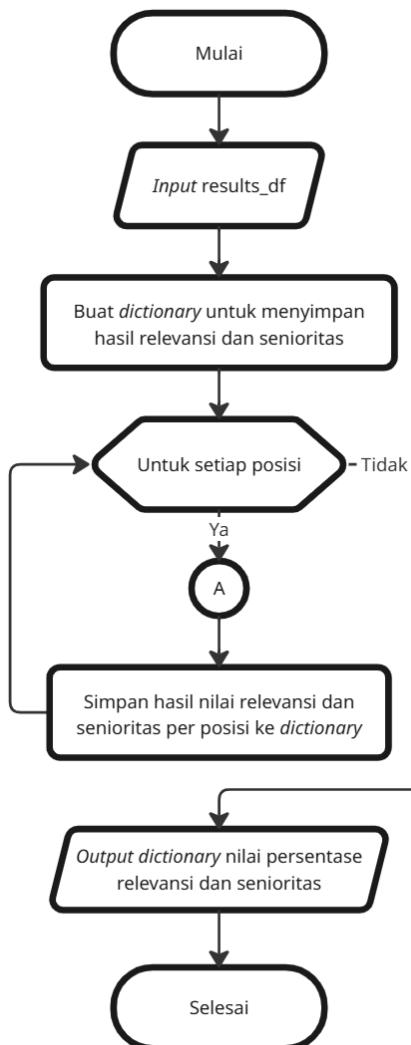
Gambar 4.19 Diagram alur *Spearman Rank Correlation Coefficient*



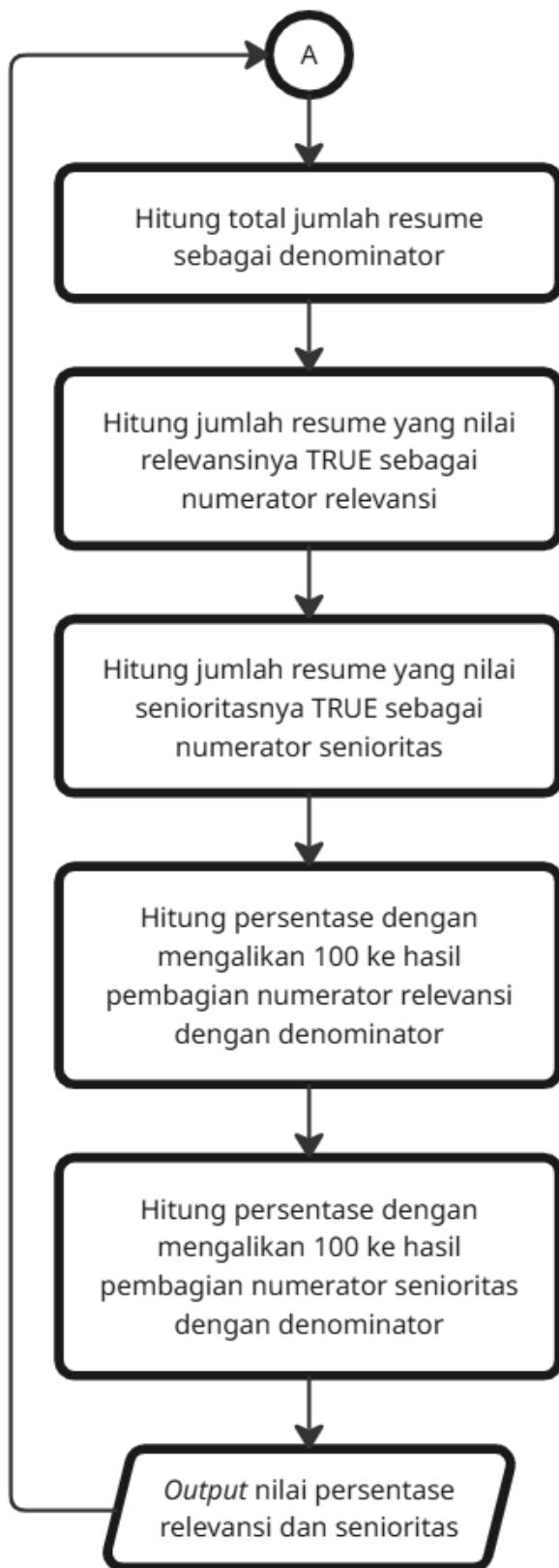
Gambar 4.20 Diagram alur *Spearman Rank Correlation Coefficient*

## 4.6 Perhitungan Relevansi dan Senioritas

Pada proses perhitungan relevansi dan senioritas, digunakan persentase untuk mengukur seberapa sesuai antara resume-resume yang dihasilkan implementasi metode dengan setiap kualifikasi lowongan kerja berdasarkan penilaian ahli. Pertama, DataFrame `results_df` yang berisi kolom peringkat hasil implementasi metode (Rank), Resume\_ID, Position, Relevance, dan Seniority diinput. Selanjutnya, untuk setiap kualifikasi lowongan kerja, hitung berapa resume yang bernilai TRUE pada kolom Relevance dan berapa resume yang bernilai TRUE pada kolom Seniority. Kemudian, hitung persentase masing-masing relevansi dan senioritas dengan mengalikan 100 pada hasil pembagian antara jumlah resume bernilai TRUE dan jumlah total resume. Diagram alur proses perhitungan relevansi dan senioritas tertera pada Gambar 4.21 hingga 4.22.



Gambar 4.21 Diagram alur relevansi dan senioritas



Gambar 4.22 Diagram alur relevansi dan senioritas

## 4.7 Perhitungan Manual

### 4.7.1 Data Uji

Data uji resume yang digunakan untuk perhitungan manualisasi merupakan salah satu resume yang diambil dari *dataset* resume kolom ID, Resume\_str, dan Resume\_html. Rincian isi data uji resume tertera pada Tabel 4.1

**Tabel 4.1 Data uji resume untuk perhitungan manual**

ID	Resume_str	Resume_html
15265464	INTERVENTION SPECIALIST TEACHER OF MATH AND LANGUAGE ARTS Objective To gain the position as the resource room teacher at Howell Township Public Schools. Summary of Qualifications Demonstrated ability to design developmentally appropriate lessons and activities allowing integration of all learning styles. Highly educated in differentiated classrooms. Determined to maximize the educational achievement of each student. Trained in Developmental Reading Assessments, Common Core Standards, Standard Solutions, Wonders, Anti-Bullying. Hard-working and organized. Knowledge and respect for all students and parental rights. Professional leadership and management skills...	<div class="fontsize fontface vmargins hmargins linespacing pagesize" id="document"> <div class="section firstsection" id="SECTION_NAME537133600" style="padding-top:5px;"> <div class="paragraph PARAGRAPH_NAME firstparagraph" id="PARAGRAPH_537133600_1_3 50680747" style="padding- top:0px;"> <div class="name thinbottomborder" itemprop="name"> <span class="field" id="537133600FNAM1"> </span> <span> </span> <span class="field" id="537133600LNAM1"> INTERVENTION SPECIALIST TEACHER OF MATH AND LANGUAGE ARTS</span> </div> <div class="myGap"> </div> <div class="lowerborder thinbottomborder"> </div> </div> </div> <div class="section" id="SECTION_SUMM537133602" style="padding-top:0px;"> <div class="heading"> <div class="sectiontitle thinbottomborder" id="SECTNAME_SUMM537133602 "> Objective</div> </div> <div class="paragraph firstparagraph" id="PARAGRAPH_537133602_1_3 50680754"...

Data uji kualifikasi lowongan kerja yang digunakan untuk perhitungan manualisasi merupakan salah satu kualifikasi lowongan kerja yang diambil dari *dataset* kualifikasi lowongan kerja dengan kolom Position, Company, dan Description. Rincian isi data uji kualifikasi lowongan kerja tertera pada Tabel 4.2

**Tabel 4.2 Data uji kualifikasi lowongan kerja untuk perhitungan manual**

<i>Position</i>	<i>Company</i>	<i>Description</i>
Creative Director / Manager	PT Basic Entertainment	<p>Strong background in event design, branding, and storytelling</p> <p>Ability to lead and inspire a team of creatives and event professionals</p> <p>Effectively present ideas to clients and collaborate with stakeholders</p> <p>Ability to understand the project scope and requirements as outlined by clients or stakeholders.</p> <p>Ensure that all designs comply with relevant industry standards</p> <p>Bachelor Degree of any major..</p>

#### **4.7.2 Perhitungan Manual Ekstraksi Section**

Pada proses ekstraksi bagian-bagian (*section*), diambil *class sectiontitle* dari kolom Resume\_html menggunakan *library* BeautifulSoup4. Setelah *section* setiap resume diketahui, isian dari setiap *section* diambil dari kolom Resume\_str seperti pada Tabel 4.3.

**Tabel 4.3 Hasil perhitungan manual ekstraksi section**

<i>Section</i>	<i>Text</i>
Objective	To gain the position as the resource room teacher at Howell Township Public Schools.
Summary of Qualifications	Demonstrated ability to design developmentally appropriate lessons and activities allowing integration of all learning styles. Highly educated in differentiated classrooms...

**Tabel 4.3 Hasil perhitungan manual ekstraksi *section* (lanjutan)**

<i>Section</i>	<i>Text</i>
Experience	Intervention Specialist Teacher of Math and Language Arts October 2013 to May 2014 Company Name — City , State Identified students with substantial academic difficulties through evaluation using Developmental Reading Assessments and consultation with staff members of referred students...
Education, Certifications, Endorsements	Bachelor of Arts : Psychology , December 2012 Georgian Court University — City , State GPA: Cum Laude Coursework in Psychology and Sociology Coursework in Intercultural and Group Communication...
Nicole Harrison Peters	732-513-7727 Nic_Harrison@aol.com

#### **4.7.3 Perhitungan Manual *Preprocessing* Resume**

##### **4.7.3.1 Menghapus *Email***

Data uji resume yang sudah melalui proses ekstraksi *section* dilakukan *preprocessing* mulai dari menghapus *email* seperti pada Tabel 4.4.

**Tabel 4.4 Hasil perhitungan manual *preprocessing* resume bagian menghapus *email***

<i>Section</i>	<i>Text</i>
Objective	To gain the position as the resource room teacher at Howell Township Public Schools.
Summary of Qualifications	Demonstrated ability to design developmentally appropriate lessons and activities allowing integration of all learning styles. Highly educated in differentiated classrooms...
Experience	Intervention Specialist Teacher of Math and Language Arts October 2013 to May 2014 Company Name — City , State Identified students with substantial academic difficulties through evaluation using Developmental Reading Assessments and consultation with staff members of referred students...

**Tabel 4.4 Hasil perhitungan manual *preprocessing* resume bagian menghapus *email* (lanjutan)**

<b>Section</b>	<b>Text</b>
Education, Certifications, Endorsements	Bachelor of Arts : Psychology , December 2012 Georgian Court University — City , State GPA: Cum Laude Coursework in Psychology and Sociology Coursework in Intercultural and Group Communication...
Nicole Harrison Peters	732-513-7727

#### **4.7.3.2 Menghapus Nomor Telepon**

Data uji resume yang sudah melalui *preprocessing* menghapus *email*, dilanjutkan menghapus nomor telepon seperti pada Tabel 4.5.

**Tabel 4.5 Hasil perhitungan manual *preprocessing* resume bagian menghapus nomor telepon**

<b>Section</b>	<b>Text</b>
Objective	To gain the position as the resource room teacher at Howell Township Public Schools.
Summary of Qualifications	Demonstrated ability to design developmentally appropriate lessons and activities allowing integration of all learning styles. Highly educated in differentiated classrooms...
Experience	Intervention Specialist Teacher of Math and Language Arts October 2013 to May 2014 Company Name — City , State Identified students with substantial academic difficulties through evaluation using Developmental Reading Assessments and consultation with staff members of referred students...
Education, Certifications, Endorsements	Bachelor of Arts : Psychology , December 2012 Georgian Court University — City , State GPA: Cum Laude Coursework in Psychology and Sociology Coursework in Intercultural and Group Communication...
Nicole Harrison Peters	

#### **4.7.3.3 Menghapus Berbagai Tipe Tanda Minus (–, –, –)**

Data uji resume yang sudah melalui *preprocessing* menghapus nomor telepon, dilanjutkan menghapus berbagai tipe tanda minus yang kemungkinan digunakan dalam penulisan informasi seperti lokasi, tanggal, atau rentang waktu. Hasil dari proses penghapusan ini tertera pada Tabel 4.6.

**Tabel 4.6 Hasil perhitungan manual *preprocessing* resume bagian menghapus berbagai tipe tanda minus (–, –, –)**

<b>Section</b>	<b>Text</b>
Objective	To gain the position as the resource room teacher at Howell Township Public Schools.
Summary of Qualifications	Demonstrated ability to design developmentally appropriate lessons and activities allowing integration of all learning styles. Highly educated in differentiated classrooms...
Experience	Intervention Specialist Teacher of Math and Language Arts October 2013 to May 2014 Company Name City , State Identified students with substantial academic difficulties through evaluation using Developmental Reading Assessments and consultation with staff members of referred students...
Education, Certifications, Endorsements	Bachelor of Arts : Psychology , December 2012 Georgian Court University City , State GPA: Cum Laude Coursework in Psychology and Sociology Coursework in Intercultural and Group Communication...
Nicole Harrison Peters	

#### **4.7.3.4 Menghapus Nama Bulan**

Data uji resume yang sudah melalui *preprocessing* menghapus berbagai tipe tanda minus, dilanjutkan menghapus nama bulan. Hasil dari proses penghapusan ini tertera pada Tabel 4.7.

**Tabel 4.7 Hasil perhitungan manual *preprocessing* resume bagian menghapus nama bulan**

<b>Section</b>	<b>Text</b>
Objective	To gain the position as the resource room teacher at Howell Township Public Schools.

**Tabel 4.7 Hasil perhitungan manual *preprocessing* resume bagian menghapus nama bulan (lanjutan)**

<b>Section</b>	<b>Text</b>
Summary of Qualifications	Demonstrated ability to design developmentally appropriate lessons and activities allowing integration of all learning styles. Highly educated in differentiated classrooms...
Experience	Intervention Specialist Teacher of Math and Language Arts 2013 to 2014 Company Name City , State Identified students with substantial academic difficulties through evaluation using Developmental Reading Assessments and consultation with staff members of referred students...
Education, Certifications, Endorsements	Bachelor of Arts : Psychology , 2012 Georgian Court University City , State GPA: Cum Laude Coursework in Psychology and Sociology Coursework in Intercultural and Group Communication...
Nicole Harrison Peters	

#### **4.7.3.5 Menghapus Kata “Present” dan “Current”**

Data uji resume yang sudah melalui *preprocessing* menghapus nama bulan, dilanjutkan menghapus kata atau istilah seperti “Present” dan “Current” yang biasanya digunakan untuk menunjukkan rentang waktu. Hasil dari proses penghapusan ini tertera pada Tabel 4.8.

**Tabel 4.8 Hasil perhitungan manual *preprocessing* resume bagian menghapus kata “Present” dan “Current”**

<b>Section</b>	<b>Text</b>
Objective	To gain the position as the resource room teacher at Howell Township Public Schools.
Summary of Qualifications	Demonstrated ability to design developmentally appropriate lessons and activities allowing integration of all learning styles. Highly educated in differentiated classrooms...

**Tabel 4.8 Hasil perhitungan manual *preprocessing* resume bagian menghapus kata “Present” dan “Current” (lanjutan)**

<b>Section</b>	<b>Text</b>
Experience	Intervention Specialist Teacher of Math and Language Arts 2013 to 2014 Company Name City , State Identified students with substantial academic difficulties through evaluation using Developmental Reading Assessments and consultation with staff members of referred students...
Education, Certifications, Endorsements	Bachelor of Arts : Psychology , 2012 Georgian Court University City , State GPA: Cum Laude Coursework in Psychology and Sociology Coursework in Intercultural and Group Communication...
Nicole Harrison Peters	

#### **4.7.3.6 Menghapus Tanggal**

Data uji resume yang sudah melalui *preprocessing* menghapus kata atau istilah seperti “Present” dan “Current”, dilanjutkan dengan penghapusan format-format tanggal dan rentang waktu. Hasil dari proses penghapusan ini tertera pada Tabel 4.9.

**Tabel 4.9 Hasil perhitungan manual *preprocessing* resume bagian menghapus tanggal**

<b>Section</b>	<b>Text</b>
Objective	To gain the position as the resource room teacher at Howell Township Public Schools.
Summary of Qualifications	Demonstrated ability to design developmentally appropriate lessons and activities allowing integration of all learning styles. Highly educated in differentiated classrooms...

**Tabel 4.9 Hasil perhitungan manual *preprocessing* resume bagian menghapus tanggal (lanjutan)**

<b>Section</b>	<b>Text</b>
Experience	Intervention Specialist Teacher of Math and Language Arts to Company Name City , State Identified students with substantial academic difficulties through evaluation using Developmental Reading Assessments and consultation with staff members of referred students...
Education, Certifications, Endorsements	Bachelor of Arts : Psychology , Georgian Court University City , State GPA: Cum Laude Coursework in Psychology and Sociology Coursework in Intercultural and Group Communication...
Nicole Harrison Peters	

#### **4.7.3.7 Menghapus Placeholder**

Data uji resume yang sudah melalui *preprocessing* menghapus tanggal, dilanjutkan menghapus *placeholder* yang biasanya digunakan untuk menunjukkan lokasi atau perusahaan seperti “Company Name” dan “State”. Hasil dari proses penghapusan ini tertera pada Tabel 4.10.

**Tabel 4.10 Hasil perhitungan manual *preprocessing* resume bagian menghapus placeholder**

<b>Section</b>	<b>Text</b>
Objective	To gain the position as the resource room teacher at Howell Township Public Schools.
Summary of Qualifications	Demonstrated ability to design developmentally appropriate lessons and activities allowing integration of all learning styles. Highly educated in differentiated classrooms...

**Tabel 4.10 Hasil perhitungan manual *preprocessing* resume bagian menghapus placeholder (lanjutan)**

<b>Section</b>	<b>Text</b>
Experience	Intervention Specialist Teacher of Math and Language Arts to , Identified students with substantial academic difficulties through evaluation using Developmental Reading Assessments and consultation with staff members of referred students...
Education, Certifications, Endorsements	Bachelor of Arts : Psychology , Georgian Court University , GPA: Cum Laude Coursework in Psychology and Sociology Coursework in Intercultural and Group Communication...
Nicole Harrison Peters	

#### **4.7.3.8 Menghapus Tanda Baca**

Data uji resume yang sudah melalui *preprocessing* menghapus *placeholder*, dilanjutkan menghapus tanda baca. Hasil dari proses penghapusan ini tertera pada Tabel 4.11.

**Tabel 4.11 Hasil perhitungan manual *preprocessing* resume bagian menghapus tanda baca**

<b>Section</b>	<b>Text</b>
Objective	To gain the position as the resource room teacher at Howell Township Public Schools
Summary of Qualifications	Demonstrated ability to design developmentally appropriate lessons and activities allowing integration of all learning styles Highly educated in differentiated classrooms...
Experience	Intervention Specialist Teacher of Math and Language Arts to Identified students with substantial academic difficulties through evaluation using Developmental Reading Assessments and consultation with staff members of referred students...

**Tabel 4.11 Hasil perhitungan manual *preprocessing* resume bagian menghapus tanda baca (lanjutan)**

<b>Section</b>	<b>Text</b>
Education, Certifications, Endorsements	Bachelor of Arts Psychology Georgian Court University GPA: Cum Laude Coursework in Psychology and Sociology Coursework in Intercultural and Group Communication...
Nicole Harrison Peters	

#### **4.7.3.9 Menghapus Angka**

Data uji resume yang sudah melalui *preprocessing* menghapus tanda baca, dilanjutkan menghapus angka. Hasil dari proses penghapusan ini tertera pada Tabel 4.12.

**Tabel 4.12 Hasil Perhitungan manual *preprocessing* resume bagian menghapus angka**

<b>Section</b>	<b>Text</b>
Objective	To gain the position as the resource room teacher at Howell Township Public Schools.
Summary of Qualifications	Demonstrated ability to design developmentally appropriate lessons and activities allowing integration of all learning styles Highly educated in differentiated classrooms...
Experience	Intervention Specialist Teacher of Math and Language Arts to Identified students with substantial academic difficulties through evaluation using Developmental Reading Assessments and consultation with staff members of referred students...
Education, Certifications, Endorsements	Bachelor of Arts Psychology Georgian Court University GPA: Cum Laude Coursework in Psychology and Sociology Coursework in Intercultural and Group Communication...
Nicole Harrison Peters	

#### **4.7.3.10 Menghapus Spasi Kosong Berlebih**

Data uji resume yang sudah melalui *preprocessing* menghapus angka, dilanjutkan menghapus spasi kosong berlebih yang biasanya muncul akibat penghapusan karakter. Hasil dari proses penghapusan ini tertera pada Tabel 4.13.

**Tabel 4.13 Hasil perhitungan manual *preprocessing* resume bagian menghapus spasi kosong berlebih**

<b>Section</b>	<b>Text</b>
Objective	To gain the position as the resource room teacher at Howell Township Public Schools
Summary of Qualifications	Demonstrated ability to design developmentally appropriate lessons and activities allowing integration of all learning styles Highly educated in differentiated classrooms...
Experience	Intervention Specialist Teacher of Math and Language Arts to Identified students with substantial academic difficulties through evaluation using Developmental Reading Assessments and consultation with staff members of referred students...
Education, Certifications, Endorsements	Bachelor of Arts Psychology Georgian Court University GPA Cum Laude Coursework in Psychology and Sociology Coursework in Intercultural and Group Communication...
Nicole Harrison Peters	

#### **4.7.3.11 Lematisasi dan Menghapus *Stop Words***

Data uji resume yang sudah melalui *preprocessing* menghapus spasi kosong berlebih, dilanjutkan menghapus *stop word* dari Bahasa Inggris menggunakan daftar *stop words* yang tersedia pada *library* Natural Language Toolkit (NLTK). Hasil dari proses penghapusan ini tertera pada Tabel 4.14.

**Tabel 4.14 Hasil perhitungan manual *preprocessing* resume bagian lematisasi dan menghapus *stop words***

<b>Section</b>	<b>Text</b>
Objective	gain position resource room teacher Howell Township Public Schools

**Tabel 4.14 Hasil perhitungan manual *preprocessing* resume bagian lematisasi dan menghapus *stop words* (lanjutan)**

<b>Section</b>	<b>Text</b>
Summary of Qualifications	Demonstrated ability design developmentally appropriate lesson activity allow integration learning style Highly educate differentiated classroom...
Experience	Intervention Specialist Teacher Math Language Arts Identified student substantial academic difficulty evaluation use Developmental Reading Assessments consultation staff member referred student...
Education, Certifications, Endorsements	Bachelor Arts Psychology Georgian Court University GPA Cum Laude Coursework Psychology Sociology Coursework Intercultural Group Communication...
Nicole Harrison Peters	

#### **4.7.3.12 Mengonversi Nama *Section* Menjadi Huruf Kecil (*Lower Casing*)**

Data uji resume yang sudah melalui *preprocessing* menghapus tanda koma berlebih, dilanjutkan mengonversi nama-nama *section* di kolom Section menjadi huruf kecil untuk memudahkan proses selanjutnya dalam merapikan nama-nama *section*. Hasil dari proses pengonversian ini tertera pada Tabel 4.15.

**Tabel 4.15 Hasil perhitungan manual penyetaraan nama *section* bagian mengonversi nama *section* menjadi huruf kecil (*lower casing*)**

<b>Section</b>	<b>Text</b>
objective	gain position resource room teacher Howell Township Public Schools
summary of qualifications	Demonstrated ability design developmentally appropriate lesson activity allow integration learning style Highly educate differentiated classroom...
experience	Intervention Specialist Teacher Math Language Arts Identified student substantial academic difficulty evaluation use Developmental Reading Assessments consultation staff member referred student...

**Tabel 4.15 Hasil perhitungan manual penyetaraan nama *section* bagian mengonversi nama *section* menjadi huruf kecil (*lower casing*) (lanjutan)**

<i>Section</i>	<i>Text</i>
education, certifications, endorements	Bachelor Arts Psychology Georgian Court University GPA Cum Laude Coursework Psychology Sociology Coursework Intercultural Group Communication...
nicole harrison peters	

#### **4.7.3.13 Mencari Nilai Unik *Section***

Data uji resume yang sudah melalui *preprocessing lower casing* nama-nama *section*, dilanjutkan dengan pencarian nama-nama *section* yang unik untuk mengidentifikasi variasi nama *section* yang terdapat dalam seluruh *dataset* resume. Hasil dari proses pencarian nama-nama *section* ini tertera pada Tabel 4.16.

**Tabel 4.16 Hasil perhitungan manual penyetaraan nama *section* bagian mengonversi nama *section* menjadi huruf kecil (*lower casing*)**

<b>Nama-Nama <i>Section</i> Unik</b>
summary
highlights
accomplishments
experience
education
...
mpd projects/clients
technical projects
core competencies

#### **4.7.3.14 Standarisasi Nama *Section***

Setelah mengetahui variasi nama *section* yang terdapat dalam seluruh *dataset* resume, dilakukan pemetaan nama-nama *section* untuk menyeragamkan agar konsisten dengan nama-nama *section* yang telah ditentukan pada penelitian ini, seperti "Summary", "Accomplishments/Awards", "Skills/Qualifications", "Education", "Experience", "Organization", "Projects", "Certifications", "Portfolio", "Others". Hasil dari proses penyetaraan nama-nama *section* ini untuk data uji resume tertera pada Tabel 4.17.

**Tabel 4.17 hasil perhitungan manual penyetaraan nama *section* bagian menyeragamkan pengelompokan *section* berdasarkan pemetaan**

<b>Section</b>	<b>Text</b>
Education	Bachelor Arts Psychology Georgian Court University GPA Cum Laude Coursework Psychology Sociology Coursework Intercultural Group Communication...
Experience	Intervention Specialist Teacher Math Language Arts Identified student substantial academic difficulty evaluation use Developmental Reading Assessments consultation staff member referred student...
Summary	gain position resource room teacher Howell Township Public Schools Demonstrated ability design developmentally appropriate lesson activity allow integration learning style Highly educate differentiated classroom...
nicole harrison peters	

#### **4.7.3.15 Klasifikasi, Penghapusan, dan Pengelompokan *Section* Tidak Valid**

Setelah penyeragaman nama-nama *section*, dilakukan proses klasifikasi, penghapusan, dan pengelompokan *section* yang tidak valid. Proses ini diawali dengan mengidentifikasi *section* yang tidak termasuk dalam daftar nama *section* yang telah ditentukan. Jika ditemukan kata kunci tertentu dalam teks, seperti “years”, “experience”, atau “I am”, maka *section* tersebut diklasifikasikan ke dalam kategori “Summary”. Jika terdapat kata “LinkedIn”, maka dikategorikan ke dalam “Portfolio”. Baris data yang memiliki isian kolom *Section* tidak sesuai dengan daftar nama *section* yang sudah ditentukan dan kolom *Text* tidak ada isinya, maka dihapus. Sedangkan, isian kolom *Section* yang tidak termasuk dalam daftar nama *section*, tetapi memiliki isi di dalam kolom *Text*, maka dipetakan ke dalam *section* “Others”. Selain itu, seluruh baris yang isian kolom *Text*-nya kosong, hanya berisi spasi atau bernilai *null* juga dihapus untuk menjaga kualitas data. Hasil dari proses klasifikasi, penghapusan, dan pengelompokan *section* ini untuk data uji resume tertera pada Tabel 4.18.

**Tabel 4.18 Hasil perhitungan manual penyetaraan nama *section* bagian klasifikasi, penghapusan, dan pengelompokan *section* tidak valid**

<i>Section</i>	<i>Text</i>
Education	Bachelor Arts Psychology Georgian Court University GPA Cum Laude Coursework Psychology Sociology Coursework Intercultural Group Communication...
Experience	Intervention Specialist Teacher Math Language Arts Identified student substantial academic difficulty evaluation use Developmental Reading Assessments consultation staff member referred student...
Summary	gain position resource room teacher Howell Township Public Schools Demonstrated ability design developmentally appropriate lesson activity allow integration learning style Highly educate differentiated classroom...

#### **4.7.3.16 Mengonversi Isi Resume Menjadi Huruf Kecil (*Lower Casing*)**

Setelah merapikan *section*, dilakukan pengonversian isi kolom *Text* menjadi huruf kecil untuk menyamakan format teks serta memudahkan proses representasi teks pada tahap selanjutnya. Hasil dari proses ini untuk data uji resume tertera pada Tabel 4.19.

**Tabel 4.19 Hasil perhitungan manual penyetaraan nama *section* bagian mengonversi isi resume menjadi huruf kecil (*lower casing*)**

<i>Section</i>	<i>Text</i>
Education	bachelor arts psychology georgian court university gpa cum laude coursework psychology sociology coursework intercultural group communication...
Experience	intervention specialist teacher math language arts identified student substantial academic difficulty evaluation use developmental reading assessments consultation staff member...
Summary	gain position resource room teacher howell township public schools demonstrated ability design developmentally appropriate lesson activity allow integration learning style highly educate differentiated classroom...

#### **4.7.4 Perhitungan Manual *Preprocessing* Kualifikasi Lowongan Kerja**

##### **4.7.4.1 Mengonversi Isi Kualifikasi Lowongan Kerja Menjadi Huruf Kecil (*Lower Casing*)**

Pada proses *preprocessing* kualifikasi lowongan kerja, dilakukan pengonversian isi kolom *Description* menjadi huruf kecil untuk menyamakan format teks serta memudahkan proses *preprocessing* selanjutnya. Hasil dari proses ini untuk data uji kualifikasi lowongan kerja tertera pada Tabel 4.20.

**Tabel 4.20 Hasil perhitungan manual *preprocessing* kualifikasi lowongan kerja bagian mengonversi isi kualifikasi lowongan kerja menjadi huruf kecil (*lower casing*)**

<b>Category</b>	<b>Position</b>	<b>Company</b>	<b>Description</b>
ARTS	Creative Director / Manager	PT Basic Entertainment	strong background in event design, branding, and storytelling ability to lead and inspire a team of creatives and event professionals...

##### **4.7.4.2 Menghapus Angka**

Setelah proses *lower casing* isi kolom *Description*, dilakukan penghapusan angka-angka. Hasil dari proses ini untuk data uji kualifikasi lowongan kerja tertera pada Tabel 4.21.

**Tabel 4.21 Hasil perhitungan manual *preprocessing* kualifikasi lowongan kerja bagian menghapus angka**

<b>Category</b>	<b>Position</b>	<b>Company</b>	<b>Description</b>
ARTS	Creative Director / Manager	PT Basic Entertainment	strong background in event design, branding, and storytelling ability to lead and inspire a team of creatives and event professionals...

#### **4.7.4.3 Menghapus Tanda Baca**

Setelah proses penghapusan angka, dilakukan penghapusan tanda baca. Hasil dari proses ini untuk data uji kualifikasi lowongan kerja tertera pada Tabel 4.22.

**Tabel 4.22 Hasil perhitungan manual *preprocessing* kualifikasi lowongan kerja bagian menghapus tanda baca**

<b>Category</b>	<b>Position</b>	<b>Company</b>	<b>Description</b>
ARTS	Creative Director / Manager	PT Basic Entertainment	strong background in event design branding and storytelling ability to lead and inspire a team of creatives and event professionals...

#### **4.7.4.4 Menghapus Spasi Kosong Berlebih**

Setelah proses penghapusan tanda baca, dilakukan penghapusan spasi kosong berlebih yang biasanya muncul akibat penghapusan karakter. Hasil dari proses ini untuk data uji kualifikasi lowongan kerja tertera pada Tabel 4.23.

**Tabel 4.23 Hasil perhitungan manual *preprocessing* kualifikasi lowongan kerja bagian menghapus spasi kosong berlebih**

<b>Category</b>	<b>Position</b>	<b>Company</b>	<b>Description</b>
ARTS	Creative Director / Manager	PT Basic Entertainment	strong background in event design branding and storytelling ability to lead and inspire a team of creatives and event professionals...

#### 4.7.4.5 Lematisasi dan Menghapus Stop Words

Setelah proses penghapusan spasi kosong berlebih, dilakukan penghapusan *stop words* dari Bahasa Inggris. Hasil dari proses ini untuk data uji kualifikasi lowongan kerja tertera pada Tabel 4.24.

**Tabel 4.24 Hasil perhitungan manual preprocessing kualifikasi lowongan kerja bagian lematisasi dan menghapus stop words**

Category	Position	Company	Description
ARTS	Creative Director / Manager	PT Basic Entertainment	strong background event design branding storytelling ability lead inspire team creatives event professional...

#### 4.7.5 Perhitungan Manual Representasi Teks

##### 4.7.5.1 Perhitungan Manual TF-IDF

Tujuan dari TF-IDF adalah menghitung tingkat kepentingan sebuah kata dalam suatu dokumen dibandingkan dengan keseluruhan dokumen dalam korpus. Untuk perhitungan manual pada resume, digunakan *sample data* dari *section “Experience”* dengan rincian tertera pada Tabel 4.25.

**Tabel 4.25 Korpus resume untuk perhitungan manual**

Dokumen 1	Dokumen 2	Dokumen 3
demonstrated ability design developmentally appropriate lesson activity allow integration learning style highly educate differentiated classroom determined maximize educational achievement student trained developmental reading assessments	digital production manager responsible aspect digital production premium printing graphic design company delivering high quality production meet client direct deadline	acted liaison senior business manager various global stake holder recruited analysts process suggest best practice effective method

Tahap pertama adalah menghitung *Term Frequency* (TF). Pada *library Scikit-learn*, nilai TF merupakan jumlah kemunculan *term* pada setiap dokumen seperti pada Persamaan 4.1.

$$TF_{(t,d)} = f_{t,d} \quad (4.1)$$

Jumlah kemunculan (frekuensi) *term* setiap dokumen untuk data uji resume tertera pada Tabel 4.26.

**Tabel 4.26 Perhitungan manual frekuensi *term* setiap resume**

<b>Term</b>	<b>f di Dokumen 1</b>	<b>f di Dokumen 2</b>	<b>f di Dokumen 3</b>
demonstrated	1	0	0
ability	1	0	0
design	1	1	0
developmentally	1	0	0
appropriate	1	0	0
lesson	1	0	0
activity	1	0	0
allow	1	0	0
integration	1	0	0
learning	1	0	0
style	1	0	0
highly	1	0	0
educate	1	0	0
differentiated	1	0	0
classroom	1	0	0
determined	1	0	0
maximize	1	0	0
educational	1	0	0
achievement	1	0	0
student	1	0	0
trained	1	0	0
developmental	1	0	0
reading	1	0	0
assessments	1	0	0
digital	0	2	0
production	0	3	0
manager	0	1	1

**Tabel 4.26 Perhitungan manual frekuensi term setiap resume (lanjutan)**

<i>Term</i>	<i>f di Dokumen 1</i>	<i>f di Dokumen 2</i>	<i>f di Dokumen 3</i>
responsible	0	1	0
aspect	0	1	0
premium	0	1	0
printing	0	1	0
graphic	0	1	0
company	0	1	0
delivering	0	1	0
high	0	1	0
quality	0	1	0
meet	0	1	0
client	0	1	0
direct	0	1	0
deadline	0	1	0
acted	0	0	1
liaison	0	0	1
senior	0	0	1
business	0	0	1
various	0	0	1
global	0	0	1
stake	0	0	1
holder	0	0	1
recruited	0	0	1
analysts	0	0	1
process	0	0	1
suggest	0	0	1
best	0	0	1
practice	0	0	1

**Tabel 4.26 Perhitungan manual frekuensi term setiap resume (lanjutan)**

Term	f di Dokumen 1	f di Dokumen 2	f di Dokumen 3
effective	0	0	1
method	0	0	1

Contoh perhitungan untuk kata “demonstrated” pada setiap dokumen tertera mulai dari Persamaan 4.2

$$TF_{(demonstrated,d1)} = 1 \quad (4.2)$$

$$TF_{(demonstrated,d2)} = 0 \quad (4.3)$$

$$TF_{(demonstrated,d3)} = 0 \quad (4.4)$$

Hasil perhitungan TF untuk seluruh *term* pada ketiga dokumen tertera pada Tabel 4.27.

**Tabel 4.27 Perhitungan manual TF korpus resume**

Term	TF di Dokumen 1	TF di Dokumen 2	TF di Dokumen 3
demonstrated	1	0	0
ability	1	0	0
design	1	1	0
developmentally	1	0	0
appropriate	1	0	0
lesson	1	0	0
activity	1	0	0
allow	1	0	0
integration	1	0	0
learning	1	0	0
style	1	0	0
highly	1	0	0
educate	1	0	0
differentiated	1	0	0
classroom	1	0	0
determined	1	0	0
maximize	1	0	0
educational	1	0	0

**Tabel 4.27 Perhitungan manual TF korpus resume (lanjutan)**

<b>Term</b>	<b>TF di Dokumen 1</b>	<b>TF di Dokumen 2</b>	<b>TF di Dokumen 3</b>
achievement	1	0	0
student	1	0	0
trained	1	0	0
developmental	1	0	0
reading	1	0	0
assessments	1	0	0
digital	0	2	0
production	0	3	0
manager	0	1	1
responsible	0	1	0
aspect	0	1	0
premium	0	1	0
printing	0	1	0
graphic	0	1	0
company	0	1	0
delivering	0	1	0
high	0	1	0
quality	0	1	0
meet	0	1	0
client	0	1	0
direct	0	1	0
deadline	0	1	0
acted	0	0	1
liaison	0	0	1
senior	0	0	1
business	0	0	1
various	0	0	1
global	0	0	1
stake	0	0	1
holder	0	0	1
recruited	0	0	1

**Tabel 4.27 Perhitungan manual TF korpus resume (lanjutan)**

Term	TF di Dokumen 1	TF di Dokumen 2	TF di Dokumen 3
analysts	0	0	1
process	0	0	1
suggest	0	0	1
best	0	0	1
practice	0	0	1
effective	0	0	1
method	0	0	1

Selanjutnya menghitung *Inverse Document Frequency* (IDF) dengan menghitung terlebih dahulu nilai *Document Frequency* (DF). Nilai DF didapatkan dari menghitung jumlah dokumen yang memiliki suatu *term*, rumusnya tertera pada Persamaan 4.5, mulai dari Persamaan 4.6 merupakan contoh perhitungan IDF untuk *term* “demonstrated”.

$$IDF_{(t)} = \log\left(\frac{1+N}{1+df}\right) + 1 \quad (4.5)$$

$$IDF_{(demonstrated)} = \log\left(\frac{1+3}{1+1}\right) + 1 = \log(2) + 1 \quad (4.6)$$

$$IDF_{(demonstrated)} = 0,6931471806 + 1 = 1.693147181 \quad (4.7)$$

Hasil IDF untuk semua *term* di korpus resume tertera pada Tabel 4.28.

**Tabel 4.28 Perhitungan manual IDF korpus resume**

Term	DF	IDF
demonstrated	1	1,693147181
ability	1	1,693147181
design	2	1,287682072
developmentally	1	1,693147181
appropriate	1	1,693147181
lesson	1	1,693147181
activity	1	1,693147181
allow	1	1,693147181
integration	1	1,693147181
learning	1	1,693147181
style	1	1,693147181
highly	1	1,693147181

**Tabel 4.28 Perhitungan manual IDF korpus resume (lanjutan)**

<b>Term</b>	<b>DF</b>	<b>IDF</b>
educate	1	1,693147181
differentiated	1	1,693147181
classroom	1	1,693147181
determined	1	1,693147181
maximize	1	1,693147181
educational	1	1,693147181
achievement	1	1,693147181
student	1	1,693147181
trained	1	1,693147181
developmental	1	1,693147181
reading	1	1,693147181
assessments	1	1,693147181
digital	1	1,693147181
production	1	1,693147181
manager	2	1,287682072
responsible	1	1,693147181
aspect	1	1,693147181
premium	1	1,693147181
printing	1	1,693147181
graphic	1	1,693147181
company	1	1,693147181
delivering	1	1,693147181
high	1	1,693147181
quality	1	1,693147181
meet	1	1,693147181
client	1	1,693147181
direct	1	1,693147181
deadline	1	1,693147181
acted	1	1,693147181
liaison	1	1,693147181
senior	1	1,693147181

**Tabel 4.28 Perhitungan manual IDF korpus resume (lanjutan)**

Term	DF	IDF
business	1	1,693147181
various	1	1,693147181
global	1	1,693147181
stake	1	1,693147181
holder	1	1,693147181
recruited	1	1,693147181
analysts	1	1,693147181
process	1	1,693147181
suggest	1	1,693147181
best	1	1,693147181
practice	1	1,693147181
effective	1	1,693147181
method	1	1,693147181

Setelah mendapatkan nilai *Term Frequency* (TF) dan *Inverse Document Frequency* (IDF), nilai TF-IDF dapat didapatkan dengan mengalikan nilai TF dengan nilai IDF seperti pada Persamaan 4.8. Contoh perhitungan manual untuk term “demonstrated” tertera mulai dari Persamaan 4.9

$$TF - IDF_{(t,d)} = TF_{(t,d)} \times IDF_{(t)} \quad (4.8)$$

$$TF - IDF_{(demonstrated,d1)} = 1 \times 1,693147181 = 1,693147181 \quad (4.9)$$

$$TF - IDF_{(demonstrated,d2)} = 0 \times 1,693147181 = 0 \quad (4.10)$$

$$TF - IDF_{(demonstrated,d3)} = 0 \times 1,693147181 = 0 \quad (4.11)$$

Hasil TF-IDF seluruh term pada ketiga dokumen tertera pada Tabel 4.29.

**Tabel 4.29 Perhitungan manual TF-IDF korpus resume**

Term	TF-IDF Dokumen 1	TF-IDF Dokumen 2	TF-IDF Dokumen 3
demonstrated	1,693147181	0	0
ability	1,693147181	0	0
design	1,287682072	1,287682072	0
developmentally	1,693147181	0	0
appropriate	1,693147181	0	0
lesson	1,693147181	0	0

**Tabel 4.29 Perhitungan manual TF-IDF korpus resume (lanjutan)**

<i>Term</i>	TF-IDF Dokumen 1	TF-IDF Dokumen 2	TF-IDF Dokumen 3
activity	1,693147181	0	0
allow	1,693147181	0	0
integration	1,693147181	0	0
learning	1,693147181	0	0
style	1,693147181	0	0
highly	1,693147181	0	0
educate	1,693147181	0	0
differentiated	1,693147181	0	0
classroom	1,693147181	0	0
determined	1,693147181	0	0
maximize	1,693147181	0	0
educational	1,693147181	0	0
achievement	1,693147181	0	0
student	1,693147181	0	0
trained	1,693147181	0	0
developmental	1,693147181	0	0
reading	1,693147181	0	0
assessments	1,693147181	0	0
digital	0	3,386294361	0
production	0	5,079441542	0
manager	0	1,287682072	1,287682072
responsible	0	1,693147181	0
aspect	0	1,693147181	0
premium	0	1,693147181	0
printing	0	1,693147181	0
graphic	0	1,693147181	0
company	0	1,693147181	0
delivering	0	1,693147181	0
high	0	1,693147181	0
quality	0	1,693147181	0
meet	0	1,693147181	0

**Tabel 4.29 Perhitungan manual TF-IDF korpus resume (lanjutan)**

<i>Term</i>	TF-IDF Dokumen 1	TF-IDF Dokumen 2	TF-IDF Dokumen 3
client	0	1,693147181	0
direct	0	1,693147181	0
deadline	0	1,693147181	0
acted	0	0	1,693147181
liaison	0	0	1,693147181
senior	0	0	1,693147181
business	0	0	1,693147181
various	0	0	1,693147181
global	0	0	1,693147181
stake	0	0	1,693147181
holder	0	0	1,693147181
recruited	0	0	1,693147181
analysts	0	0	1,693147181
process	0	0	1,693147181
suggest	0	0	1,693147181
best	0	0	1,693147181
practice	0	0	1,693147181
effective	0	0	1,693147181
method	0	0	1,693147181

Secara bawaan, TfIdfTransformer dari *library* Scikit-learn mengimplementasikan  $L_2$  (*Euclidean distance*) norm pada hasil perhitungan TF-IDF. Perhitungan normalisasi tertera pada Persamaan 4.12 dan contoh perhitungannya untuk term “*demonstrated*” tertera mulai dari Persamaan 4.13

$$v_{norm} = \frac{v}{\|v\|_2} = \frac{v}{\sqrt{v_1^2 + v_2^2 + \dots + v_n^2}} \quad (4.12)$$

$$v_{norm(demonstrated,d1)} = \frac{1,693147181}{\sqrt{(1,693147181)^2 + (1,693147181)^2 + \dots + (0)^2}} \quad (4.13)$$

$$v_{norm(demonstrated,d1)} = \frac{1,693147181}{8,221515356} = 0,2059410105 \quad (4.14)$$

$$v_{norm(demonstrated,d2)} = \frac{0}{\sqrt{(0)^2 + (0)^2 + \dots + (0)^2}} \quad (4.15)$$

$$v_{norm(demonstrated,d2)} = \frac{0}{8,823360017} = 0 \quad (4.16)$$

$$v_{norm(demonstrated,d3)} = \frac{0}{\sqrt{(0)^2 + (0)^2 + \dots + (1,693147181)^2}} \quad (4.17)$$

$$v_{norm(demonstrated,d3)} = \frac{0}{6,893916385} = 0 \quad (4.18)$$

Hasil akhir TF-IDF setelah normalisasi tertera pada tabel 4.30.

**Tabel 4.30 Perhitungan manual TF-IDF korpus resume setelah normalisasi**

Term	TF-IDF Dokumen 1	TF-IDF Dokumen 2	TF-IDF Dokumen 3
demonstrated	0,2059410105	0	0
ability	0,2059410105	0	0
design	0,1566234467	0,1459401033	0
developmentally	0,2059410105	0	0
appropriate	0,2059410105	0	0
lesson	0,2059410105	0	0
activity	0,2059410105	0	0
allow	0,2059410105	0	0
integration	0,2059410105	0	0
learning	0,2059410105	0	0
style	0,2059410105	0	0
highly	0,2059410105	0	0
educate	0,2059410105	0	0
differentiated	0,2059410105	0	0
classroom	0,2059410105	0	0
determined	0,2059410105	0	0
maximize	0,2059410105	0	0

**Tabel 4.30 Perhitungan manual TF-IDF korpus resume setelah normalisasi  
(lanjutan)**

Term	TF-IDF Dokumen 1	TF-IDF Dokumen 2	TF-IDF Dokumen 3
educational	0,2059410105	0	0
achievement	0,2059410105	0	0
student	0,2059410105	0	0
trained	0,2059410105	0	0
developmental	0,2059410105	0	0
reading	0,2059410105	0	0
assessments	0,2059410105	0	0
digital	0	0,3837873956	0
production	0	0,5756810934	0
manager	0	0,1459401033	0,1867852757
responsible	0	0,1918936978	0
aspect	0	0,1918936978	0
premium	0	0,1918936978	0
printing	0	0,1918936978	0
graphic	0	0,1918936978	0
company	0	0,1918936978	0
delivering	0	0,1918936978	0
high	0	0,1918936978	0
quality	0	0,1918936978	0
meet	0	0,1918936978	0
client	0	0,1918936978	0
direct	0	0,1918936978	0
deadline	0	0,1918936978	0
acted	0	0	0,245600191
liaison	0	0	0,245600191
senior	0	0	0,245600191
business	0	0	0,245600191
various	0	0	0,245600191
global	0	0	0,245600191
stake	0	0	0,245600191

**Tabel 4.30 Perhitungan manual TF-IDF korpus resume setelah normalisasi (lanjutan)**

Term	TF-IDF Dokumen 1	TF-IDF Dokumen 2	TF-IDF Dokumen 3
holder	0	0	0,245600191
recruited	0	0	0,245600191
analysts	0	0	0,245600191
process	0	0	0,245600191
suggest	0	0	0,245600191
best	0	0	0,245600191
practice	0	0	0,245600191
effective	0	0	0,245600191
method	0	0	0,245600191

Perhitungan manual pada kualifikasi lowongan kerja menggunakan beberapa *sample data* dari kolom *Description* dengan rincian yang bisa dilihat pada Tabel 4.31.

**Tabel 4.31 Korpus kualifikasi lowongan kerja untuk perhitungan manual**

Dokumen 4	Dokumen 5	Dokumen 6
strong background event design branding storytelling ability lead inspire team creatives event professional effectively present idea client collaborate stakeholder	execute daytoday design request internal external communication material develop creative visual content include motion graphic static design	play vital role support lead teacher create nurturing stimulate learning environment young student assist lead teacher planning implement engage educational activity

Nilai *Term Frequency* (TF) untuk deskripsi kualifikasi lowongan kerja dapat dihitung berdasarkan *term* yang telah ditokenisasi sebelumnya dari korpus resume, sehingga hasil perhitungan TF kualifikasi lowongan kerja tertera pada Tabel 4.32

**Tabel 4.32 Perhitungan manual TF korpus kualifikasi lowongan kerja**

Term	TF di Dokumen 4	TF di Dokumen 5	TF di Dokumen 6
demonstrated	0	0	0
ability	1	0	0
design	1	2	0

**Tabel 4.32 Perhitungan manual TF korpus kualifikasi lowongan kerja (lanjutan)**

<i>Term</i>	TF di Dokumen 4	TF di Dokumen 5	TF di Dokumen 6
developmentally	0	0	0
appropriate	0	0	0
lesson	0	0	0
activity	0	0	1
allow	0	0	0
integration	0	0	0
learning	0	0	1
style	0	0	0
highly	0	0	0
educate	0	0	0
differentiated	0	0	0
classroom	0	0	0
determined	0	0	0
maximize	0	0	0
educational	0	0	1
achievement	0	0	0
student	0	0	1
trained	0	0	0
developmental	0	0	0
reading	0	0	0
assessments	0	0	0
digital	0	0	0
production	0	0	0
manager	0	0	0
responsible	0	0	0
aspect	0	0	0
premium	0	0	0
printing	0	0	0

**Tabel 4.32 Perhitungan manual TF korpus kualifikasi lowongan kerja (lanjutan)**

<i>Term</i>	TF di Dokumen 4	TF di Dokumen 5	TF di Dokumen 6
graphic	0	1	0
company	0	0	0
delivering	0	0	0
high	0	0	0
quality	0	0	0
meet	0	0	0
client	1	0	0
direct	0	0	0
deadline	0	0	0
acted	0	0	0
liaison	0	0	0
senior	0	0	0
business	0	0	0
various	0	0	0
global	0	0	0
stake	0	0	0
holder	0	0	0
recruited	0	0	0
analysts	0	0	0
process	0	0	0
suggest	0	0	0
best	0	0	0
practice	0	0	0
effective	0	0	0
method	0	0	0

Nilai *Inverse Document Frequency* (IDF) bisa didapatkan dengan menghitung *document frequency* (DF) terlebih dahulu. Kemudian, dihitung dengan formula seperti pada Persamaan 4.19. Mulai dari Persamaan 4.20 mendemonstrasikan perhitungan IDF untuk term "design".

$$IDF_{(t)} = \log\left(\frac{1+N}{1+df}\right) + 1 \quad (4.19)$$

$$IDF_{(design)} = \log\left(\frac{1+3}{1+2}\right) + 1 = \log\left(\frac{4}{3}\right) + 1 \quad (4.20)$$

$$IDF_{(design)} = 0,2876820725 + 1 = 1,287682072 \quad (4.21)$$

Hasil perhitungan IDF korpus kualifikasi lowongan kerja tertera pada Tabel 4.33.

**Tabel 4.33 Perhitungan manual IDF korpus kualifikasi lowongan kerja**

Term	DF	IDF
demonstrated	0	2,386294361
ability	1	1,693147181
design	2	1,287682072
developmentally	0	2,386294361
appropriate	0	2,386294361
lesson	0	2,386294361
activity	1	1,693147181
allow	0	2,386294361
integration	0	2,386294361
learning	1	1,693147181
style	0	2,386294361
highly	0	2,386294361
educate	0	2,386294361
differentiated	0	2,386294361
classroom	0	2,386294361
determined	0	2,386294361
maximize	0	2,386294361
educational	1	1,693147181
achievement	0	2,386294361
student	1	1,693147181
trained	0	2,386294361
developmental	0	2,386294361

**Tabel 4.33 Perhitungan manual IDF korpus kualifikasi lowongan kerja (lanjutan)**

<i>Term</i>	<i>DF</i>	<i>IDF</i>
reading	0	2,386294361
assessments	0	2,386294361
digital	0	2,386294361
production	0	2,386294361
manager	0	2,386294361
responsible	0	2,386294361
aspect	0	2,386294361
premium	0	2,386294361
printing	0	2,386294361
graphic	1	1,693147181
company	0	2,386294361
delivering	0	2,386294361
high	0	2,386294361
quality	0	2,386294361
meet	0	2,386294361
client	1	1,693147181
direct	0	2,386294361
deadline	0	2,386294361
acted	0	2,386294361
liaison	0	2,386294361
senior	0	2,386294361
business	0	2,386294361
various	0	2,386294361
global	0	2,386294361
stake	0	2,386294361
holder	0	2,386294361
recruited	0	2,386294361
analysts	0	2,386294361
process	0	2,386294361
suggest	0	2,386294361
best	0	2,386294361

**Tabel 4.33 Perhitungan manual IDF korpus kualifikasi lowongan kerja (lanjutan)**

<i>Term</i>	<i>DF</i>	<i>IDF</i>
practice	0	2,386294361
effective	0	2,386294361
method	0	2,386294361

Setelah mendapatkan nilai *term frequency* (TF) dan *inverse document frequency* (IDF), TF-IDF untuk kualifikasi lowongan kerja dihitung menggunakan formula pada Persamaan 4.22. Mulai dari Persamaan 4.23 merupakan demonstrasi perhitungan TF-IDF untuk term “*design*”.

$$TF - IDF_{(t,d)} = TF_{(t,d)} \times IDF_{(t)} \quad (4.22)$$

$$TF - IDF_{(design,d4)} = 1 \times 1,287682072 = 1,287682072 \quad (4.23)$$

$$TF - IDF_{(design,d5)} = 2 \times 1,287682072 = 2,575364145 \quad (4.24)$$

$$TF - IDF_{(design,d6)} = 0 \times 1,287682072 = 0 \quad (4.25)$$

Hasil perhitungan manual TF-IDF korpus kualifikasi lowongan kerja tertera pada Tabel 4.34.

**Tabel 4.34 Perhitungan manual TF-IDF korpus kualifikasi lowongan kerja**

<i>Term</i>	<i>TF-IDF Dokumen 4</i>	<i>TF-IDF Dokumen 5</i>	<i>TF-IDF Dokumen 6</i>
demonstrated	0	0	0
ability	1,693147181	0	0
design	1,287682072	2,575364145	0
developmentally	0	0	0
appropriate	0	0	0
lesson	0	0	0
activity	0	0	1,693147181
allow	0	0	0
integration	0	0	0
learning	0	0	1,693147181
style	0	0	0
highly	0	0	0
educate	0	0	0
differentiated	0	0	0
classroom	0	0	0
determined	0	0	0

**Tabel 4.34 Perhitungan manual TF-IDF korpus kualifikasi lowongan kerja (lanjutan)**

Term	TF-IDF Dokumen 4	TF-IDF Dokumen 5	TF-IDF Dokumen 6
maximize	0	0	0
educational	0	0	1,693147181
achievement	0	0	0
student	0	0	1,693147181
trained	0	0	0
developmental	0	0	0
reading	0	0	0
assessments	0	0	0
digital	0	0	0
production	0	0	0
manager	0	0	0
responsible	0	0	0
aspect	0	0	0
premium	0	0	0
printing	0	0	0
graphic	0	1,693147181	0
company	0	0	0
delivering	0	0	0
high	0	0	0
quality	0	0	0
meet	0	0	0
client	1,693147181	0	0
direct	0	0	0
deadline	0	0	0
acted	0	0	0
liaison	0	0	0
senior	0	0	0
business	0	0	0
various	0	0	0
global	0	0	0

**Tabel 4.34 Perhitungan manual TF-IDF korpus kualifikasi lowongan kerja (lanjutan)**

Term	TF-IDF Dokumen 4	TF-IDF Dokumen 5	TF-IDF Dokumen 6
stake	0	0	0
holder	0	0	0
recruited	0	0	0
analysts	0	0	0
process	0	0	0
suggest	0	0	0
best	0	0	0
practice	0	0	0
effective	0	0	0
method	0	0	0

Hasil TF-IDF korpus kualifikasi lowongan kerja juga dilakukan normalisasi menggunakan formula pada Persamaan 4.26. Mulai dari Persamaan 4.27 merupakan demonstrasi perhitungan normalisasi TF-IDF untuk term “design”

$$v_{norm} = \frac{v}{\|v\|_2} = \frac{v}{\sqrt{v_1^2 + v_2^2 + \dots + v_n^2}} \quad (4.26)$$

$$v_{norm(design,d4)} = \frac{1,287682072}{\sqrt{(0)^2 + (1,693147181)^2 + \dots + (0)^2}} \quad (4.27)$$

$$v_{norm(design,d4)} = \frac{1,287682072}{2,718753367} = 0,473629601 \quad (4.28)$$

$$v_{norm(design,d5)} = \frac{2,575364145}{\sqrt{(0)^2 + (0)^2 + \dots + (0)^2}} \quad (4.29)$$

$$v_{norm(design,d5)} = \frac{2,575364145}{3,082084985} = 0,8355915419 \quad (4.30)$$

$$v_{norm(design,d6)} = \frac{0}{\sqrt{(0)^2 + (0)^2 + \dots + (0)^2}} \quad (4.31)$$

$$v_{norm(design,d6)} = \frac{0}{3,386294361} = 0 \quad (4.32)$$

Hasil perhitungan manual TF-IDF korpus kualifikasi lowongan kerja setelah normalisasi tertera pada Tabel 4.35.

**Tabel 4.35 Perhitungan manual TF-IDF korpus kualifikasi lowongan kerja setelah normalisasi**

Term	TF-IDF Dokumen 4	TF-IDF Dokumen 5	TF-IDF Dokumen 6
demonstrated	0	0	0
ability	0,6227660078	0	0
design	0,473629601	0,8355915419	0

**Tabel 4.35 Perhitungan manual TF-IDF korpus kualifikasi lowongan kerja setelah normalisasi (lanjutan)**

Term	TF-IDF Dokumen 4	TF-IDF Dokumen 5	TF-IDF Dokumen 6
developmentally	0	0	0
appropriate	0	0	0
lesson	0	0	0
activity	0	0	0,5
allow	0	0	0
integration	0	0	0
learning	0	0	0,5
style	0	0	0
highly	0	0	0
educate	0	0	0
differentiated	0	0	0
classroom	0	0	0
determined	0	0	0
maximize	0	0	0
educational	0	0	0,5
achievement	0	0	0
student	0	0	0,5
trained	0	0	0
developmental	0	0	0
reading	0	0	0
assessments	0	0	0
digital	0	0	0
production	0	0	0
manager	0	0	0
responsible	0	0	0
aspect	0	0	0
premium	0	0	0
printing	0	0	0
graphic	0	0,549351231	0
company	0	0	0

**Tabel 4.35 Perhitungan manual TF-IDF korpus kualifikasi lowongan kerja setelah normalisasi (lanjutan)**

Term	TF-IDF Dokumen 4	TF-IDF Dokumen 5	TF-IDF Dokumen 6
delivering	0	0	0
high	0	0	0
quality	0	0	0
meet	0	0	0
client	0,6227660078	0	0
direct	0	0	0
deadline	0	0	0
acted	0	0	0
liaison	0	0	0
senior	0	0	0
business	0	0	0
various	0	0	0
global	0	0	0
stake	0	0	0
holder	0	0	0
recruited	0	0	0
analysts	0	0	0
process	0	0	0
suggest	0	0	0
best	0	0	0
practice	0	0	0
effective	0	0	0
method	0	0	0

#### 4.7.5.2 Perhitungan Manual Word2Vec

Tujuan dari *Skip-gram* adalah memprediksi konteks (*output*) disekitar kata target (*input*). Untuk demonstrasi perhitungan manual Word2Vec menggunakan salah satu kalimat dari resume, yakni “*demonstrated ability design developmentally appropriate lesson activity allow integration learning style highly educate differentiated classroom determined maximize educational achievement student trained developmental reading assessments*” dengan parameter dengan *vocab* = 24, *window size* = 3, dan *learning rate* = 0,1.



Gambar 4.23 Pasangan target-konteks Word2Vec *Skip-gram*

Gambar 4.23 merupakan rincian pasangan target-konteks, kotak berwarna biru merupakan kata target dan kotak berwarna merah merupakan konteks dari kata target yang sesuai dengan nilai *window size*.

Tabel 4.36 merupakan *one-hot encoding* untuk setiap token dalam *vocabulary* [“*Demonstrated*”, “*ability*”, “*design*”, “*developmentally*”, “*appropriate*”, “*lesson*”, “*activity*”, “*integration*”, “*learning*”, “*style*”, “*differentiated*”, “*classroom*”, “*determined*”, “*maximize*”, “*educational*”, “*achievement*”, “*student*”, “*trained*” “*developmental*”, “*reading*”, “*assessments*”].

Tabel 4.36 Perhitungan manual *one-hot encoding*

Term	One-Hot Encoding
demonstrated	[1 0]
ability	[0 1 0]
design	[0 0 1 0]
developmentally	[0 0 0 1 0]
appropriate	[0 0 0 0 1 0]
lesson	[0 0 0 0 0 1 0]

**Tabel 4.36 Perhitungan manual *one-hot encoding* (lanjutan)**

<i>Term</i>	<i>One-Hot Encoding</i>
activity	[0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
allow	[0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
integration	[0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
learning	[0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0]
style	[0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0]
highly	[0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0]
educate	[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0]
differentiated	[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0]
classroom	[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0]
determined	[0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0]
maximize	[0 1 0 0 0 0 0 0 0]
educational	[0 1 0 0 0 0 0 0]
achievement	[0 1 0 0 0 0 0]
student	[0 1 0 0 0 0]
trained	[0 1 0 0 0]
developmental	[0 1 0 0]
reading	[0 1 0]
assessments	[0 1]

Pada perhitungan manual ini, dilakukan penargetan pada kata “*design*” dan menggunakan *hidden layer* = 3. Selanjutnya adalah melakukan inisialisasi acak bobot *input layer* ke tiga *hidden layer* yang tertera pada Tabel 4.37.

**Tabel 4.37 Bobot *input layer-hidden layer***

<b>Term</b>	<b>H1</b>	<b>H2</b>	<b>H3</b>
demonstrated	-0,3	-0,8	-0,3
ability	0,2	-0,5	-0,6
design	0,8	-0,6	-0,2
developmentally	-0,8	0	0,9
appropriate	-0,3	0,2	-0,8
lesson	-0,8	-0,3	0,2
activity	-0,6	0,2	-0,3
allow	0,2	-0,3	-0,8
integration	-0,3	0,2	-0,6
learning	-0,8	-0,3	0,2
style	-0,6	0,2	-0,3
highly	0,2	-0,3	-0,8
educate	-0,3	0,2	-0,6
differentiated	-0,8	-0,3	0,2
classroom	-0,6	0,2	-0,3
determined	0,2	-0,3	-0,8
maximize	-0,3	-0,5	0,2
educational	0,2	-0,3	-0,5
achievement	-0,8	0,2	-0,3
student	-0,5	-0,3	0,2
trained	-0,3	0,2	-0,8
developmental	0,2	-0,3	-0,5
reading	-0,8	0,2	-0,3
assessments	-0,5	-0,3	0,2

Setelah menginisialisasi acak bobot *input* ke *hidden layer*, dapat menginisialisasi acak bobot juga untuk *hidden layer* ke *output layer* yang tertera pada Tabel 4.38 sampai dengan Tabel 4.41.

**Tabel 4.38 Bobot *hidden layer-output layer***

<i>Neuron</i>	<i>demonstrated</i>	<i>ability</i>	<i>design</i>	<i>developmentally</i>	<i>appropriate</i>	<i>lesson</i>
H1	0,3	-0,2	-0,2	0	-0,2	-0,1
H2	-0,4	-0,3	-0,6	0,3	-0,1	-0,2
H3	-0,1	-0,8	0,5	0,2	-0,4	0,3

**Tabel 4.39 Bobot *hidden layer-output layer***

<i>Neuron</i>	<i>activity</i>	<i>allow</i>	<i>integration</i>	<i>learning</i>	<i>style</i>	<i>highly</i>
H1	-0,1	0,3	-0,2	-0,1	-0,1	0,3
H2	0,3	-0,2	0,3	-0,2	0,3	-0,2
H3	-0,2	-0,1	-0,1	0,3	-0,2	-0,1

**Tabel 4.40 Bobot *hidden layer-output layer***

<i>Neuron</i>	<i>educate</i>	<i>differentiated</i>	<i>classroom</i>	<i>determined</i>	<i>maximize</i>	<i>educational</i>
H1	-0,2	-0,1	-0,1	0,3	-0,2	-0,1
H2	0,3	-0,2	0,3	-0,2	0,3	-0,2
H3	-0,1	0,3	-0,2	-0,1	-0,1	0,3

**Tabel 4.41 Bobot *hidden layer-output layer***

<i>Neuron</i>	<i>achievement</i>	<i>student</i>	<i>trained</i>	<i>developmental</i>	<i>reading</i>	<i>assessments</i>
H1	-0,1	0,3	-0,2	-0,4	-0,1	0,3
H2	0,3	-0,2	0,3	-0,1	0,3	0
H3	-0,2	-0,1	-0,1	0,3	0	-0,6

Telah didapatkan bobot untuk *hidden layer* ke *output layer* pada Tabel 4.38 sampai dengan Tabel 4.41. Perhitungan dilanjutkan ke *feedforward* dengan mengalikan *one-hot vector* dari kata target “*design*” dengan matriks bobot *input-hidden* seperti pada Persamaan 4.33 sampai dengan Persamaan 4.39.

$$H1 = [0 \ 0 \ 1 \ \dots \ 0 \ 0] \times \begin{bmatrix} -0,3 \\ 0,2 \\ 0,8 \\ \vdots \\ -0,8 \\ -0,5 \end{bmatrix} \quad (4.33)$$

$$H1 = 0(-0,3) + 0(0,2) + 1(0,8) + \dots + 0(-0,8) + 0(-0,5) = 0,8 \quad (4.34)$$

$$H2 = [0 \ 0 \ 1 \ \dots \ 0 \ 0] \times \begin{bmatrix} -0,8 \\ -0,5 \\ -0,6 \\ \vdots \\ 0,2 \\ -0,3 \end{bmatrix} \quad (4.35)$$

$$H2 = 0(-0,8) + 0(-0,5) + 1(-0,6) + \dots + 0(0,2) + 0(-0,3) = -0,6 \quad (4.36)$$

$$H3 = [0 \ 0 \ 1 \ \dots \ 0 \ 0] \times \begin{bmatrix} -0,3 \\ -0,6 \\ -0,2 \\ \vdots \\ -0,3 \\ 0,2 \end{bmatrix} \quad (4.37)$$

$$H3 = 0(-0,3) + 0(-0,6) + 1(-0,2) + \dots + 0(-0,3) + 0(0,2) = -0,2 \quad (4.38)$$

$$h_{design} = [0,8 \ -0,6 \ -0,2] \quad (4.39)$$

Kalikan  $h_{design}$  dengan bobot *hidden-output* untuk menghitung skor prediksi kata target dengan konteksnya tertera pada Persamaan 4.40 sampai dengan Persamaan 4.41.

$$u_{1,1} = [0,8 \ -0,6 \ -0,2] \times \begin{bmatrix} 0,3 & -0,2 & -0,2 & \dots & 0,3 \\ -0,4 & -0,3 & 0,3 & \dots & 0 \\ -0,1 & -0,8 & 0,2 & \dots & -0,6 \end{bmatrix} \quad (4.40)$$

$$u_{1,1} = 0,8(0,3) + (-0,6)(-0,4) + (-0,2)(-0,1) = 0,5 \quad (4.41)$$

Matriks hasil berbentuk  $1 \times 24$  yang tertera pada Persamaan 4.42.

$$z_{design} = [0,5 \ 0,18 \ 0,1 \ -0,22 \ \dots \ -0,32 \ -0,26 \ 0,36] \quad (4.42)$$

Setelah didapatkan skor prediksi, lakukan normalisasi dengan *Softmax* dengan rumus seperti Persamaan 4.43.

$$\hat{y}_k = P_r(kata_k | kata_{konteks}) = \frac{\exp(k)}{\sum_n \exp(n)} \quad (4.43)$$

Jumlahkan semua nilai eksponensial untuk denominator seperti Persamaan 4.44 sampai dengan 4.45.

$$\exp(0,5) + \exp(0,18) + \dots + \exp(-0,26) + \exp(0,36) = 24,5489614 \quad (4.44)$$

$$u_{1,1} = \frac{1,648721271}{24,54896145} = 0,06716053035 \quad (4.45)$$

Matriks *Softmax* untuk kata target “*design*” berbentuk  $1 \times 24$  tertera pada Persamaan 4.46.

$$\hat{y}_{design} = [0,0671605 \quad 0,04876856 \quad \dots \quad 0,0314087 \quad 0,0583866] \quad (4.46)$$

Selanjutnya dilakukan *backpropagation* untuk memperbarui nilai bobot matriks dengan rumus seperti Persamaan 4.47.

$$\frac{dLoss}{dz} = dz = \hat{y} - y(1 \times V) \quad (4.47)$$

Diketahui nilai  $y$  untuk konteks “*ability*” pada Persamaan 4.48 dan “*development*” pada Persamaan 4.49.

$$y_{ability} = [0 \quad 1 \quad 0 \quad 0 \quad \dots \quad 0 \quad 0 \quad 0] \quad (4.48)$$

$$y_{development} = [0 \quad 0 \quad 0 \quad 1 \quad \dots \quad 0 \quad 0 \quad 0] \quad (4.49)$$

Maka, perhitungan *error* konteks “*ability*” pada kata target “*design*” tertera pada Persamaan 4.50 sampai dengan Persamaan 4.51.

$$e_{ability} = [0,0672 \quad 0,0488 \quad \dots \quad 0,0314 \quad 0,0584] - [0 \quad 1 \quad \dots \quad 0] \quad (4.50)$$

$$e_{ability} = [0,0672 \quad -0,9512 \quad \dots \quad 0,0314 \quad 0,0584] \quad (4.51)$$

Perhitungan *error* konteks “*development*” pada kata target “*design*” tertera pada Persamaan 4.52 sampai dengan Persamaan 4.53.

$$e_{development} = [0,0672 \quad 0,0488 \quad \dots \quad 0,0314 \quad 0,0584] - [0 \quad \dots \quad 0] \quad (4.52)$$

$$e_{development} = [0,0672 \quad 0,0488 \quad \dots \quad 0,0314 \quad 0,0584] \quad (4.53)$$

Perhitungan *error* rata-rata dari kedua konteks tertera pada Persamaan 4.54 sampai dengan Persamaan 4.56.

$$e_{context} = \frac{e_{ability} + e_{development}}{2} \quad (4.54)$$

$$e_{context_{1,1}} = \frac{0,06716053035 + 0,06716053035}{2} = 0,06716053035 \quad (4.55)$$

$$e_{context_{1,2}} = \frac{-0,9512314456 + 0,04876855444}{2} = -0,4512314456 \quad (4.56)$$

Matriks *error* konteks tertera pada Persamaan 4.57.

$$dz = [0,06716053 \quad -0,4512314 \quad \dots \quad 0,03140873 \quad 0,05838656] \quad (4.57)$$

Setelah mendapatkan hasil perhitungan *error*, dilanjutkan dengan perhitungan gradien dari *output layer* ke *hidden layer* dengan rumus seperti pada Persamaan 4.58.

$$\frac{dLoss}{dU} = dU = h^T \cdot dz(N \times V) \quad (4.58)$$

Diketahui matriks  $h_{design}$  dan dilakukan *transpose* pada Persamaan 4.59 sampai dengan 4.60.

$$h_{design} = [0,8 \quad -0,6 \quad -0,2] \quad (4.59)$$

$$h^T_{design} = \begin{bmatrix} 0,8 \\ -0,6 \\ -0,2 \end{bmatrix} \quad (4.60)$$

Maka, perhitungan  $dU$  tertera pada Persamaan 4.61.

$$dU = \begin{bmatrix} 0,8 \\ -0,6 \\ -0,2 \end{bmatrix} \times [0,0672 \quad -0,4512 \quad \dots \quad 0,0314 \quad 0,0584] \quad (4.61)$$

$$dU_{1,1} = 0,8 \times 0,067160530 = 0,05372842428 \quad (4.62)$$

$$dU_{2,1} = 0,6 \times 0,067160530 = -0,04029631821 \quad (4.63)$$

$$dU_{1,2} = 0,8 \times -0,45123144 = -0,3609851564 \quad (4.64)$$

$$dU_{2,2} = 0,6 \times -0,45123144 = 0,2707388673 \quad (4.65)$$

Matriks  $dU$  berbentuk  $3 \times 24$  tertera pada Persamaan 4.66.

$$dU = \begin{bmatrix} 0,05372842428 & -0,3609851564 & \dots & 0,04670924812 \\ -0,04029631821 & 0,2707388673 & \dots & -0,03503193609 \\ -0,01343210607 & 0,09024628911 & \dots & -0,01167731203 \end{bmatrix} \quad (4.66)$$

Setelah mendapatkan hasil perhitungan gradien dari *output layer* ke *hidden layer*, dilanjutkan menghitung gradien dari *hidden layer* ke *input layer* dengan rumus seperti pada Persamaan 4.67.ex

$$\frac{dLoss}{dh} = \frac{dLoss}{dz} \cdot \frac{dz}{dh} = dz \cdot U^T (1 \times N) \quad (4.67)$$

Diketahui matriks *hidden layer* 1, 2, dan 3 dilakukan transpose pada Persamaan 4.68 sampai dengan 4.73.

$$h_1 = [0,3 \quad -0,2 \quad -0,2 \quad \dots \quad 0,3] \quad (4.68)$$

$$h^T_1 = \begin{bmatrix} 0,2 \\ -0,2 \\ -0,2 \\ \vdots \\ 0,3 \end{bmatrix} \quad (4.69)$$

$$h_2 = [-0,4 \quad -0,3 \quad -0,6 \quad \dots \quad 0] \quad (4.70)$$

$$h^T_2 = \begin{bmatrix} -0,4 \\ -0,3 \\ -0,6 \\ \vdots \\ 0 \end{bmatrix} \quad (4.71)$$

$$h_3 = [-0,1 \quad -0,8 \quad -0,5 \quad \dots \quad -0,6] \quad (4.72)$$

$$h^T_3 = \begin{bmatrix} -0,1 \\ -0,8 \\ -0,5 \\ \vdots \\ -0,6 \end{bmatrix} \quad (4.73)$$

Maka, perhitungan  $\frac{dLoss}{dh}$  tertera pada Persamaan 4.74 sampai dengan Persamaan 4.76.

$$\frac{dLoss}{dh_1} = [0,0672 \quad -0,4512 \quad \dots \quad 0,0314 \quad 0,0584] \times \begin{bmatrix} 0,2 \\ -0,2 \\ -0,2 \\ \vdots \\ 0,3 \end{bmatrix} \quad (4.74)$$

$$\frac{dLoss}{dh_1} = (0,0672 \times 0,2) + (-0,4512 \times (-0,2)) + \dots + (0,0584 \times 0,3) \quad (4.75)$$

$$\frac{dLoss}{dh_1} = 0,1147162368 \quad (4.76)$$

Hitung dengan persamaan yang sama untuk  $\frac{dLoss}{dh_2}$  dan  $\frac{dLoss}{dh_3}$ . Maka hasil perhitungan gradien dari *hidden layer* ke *input layer* tertera pada Persamaan 4.77 sampai dengan 4.79.

$$\frac{dLoss}{dh_1} = 0,1147162368 \quad (4.77)$$

$$\frac{dLoss}{dh_2} = -0,06109851674 \quad (4.78)$$

$$\frac{dLoss}{dh_3} = 0,2272906395 \quad (4.79)$$

Pembaharuan bobot dihitung menggunakan *learning rate* dan gradien dari *hidden layer* ke *input layer* dilakukan seperti pada Persamaan 4.80 sampai dengan Persamaan 4.82.

$$W_{1,1} = 0,8 - (0,1 \times 0,1147162368) = 0,7885283763 \quad (4.80)$$

$$W_{1,2} = (-0,6) - (0,1 \times -0,06109851674) = -0,5938901483 \quad (4.81)$$

$$W_{1,3} = (-0,2) - (0,1 \times 0,2272906395) = -0,222729064 \quad (4.82)$$

Hasil pembaharuan bobot pada kata “*design*” tertera pada Tabel 4.42.

**Tabel 4.42 Pembaharuan bobot *input layer-hidden layer***

<b>Term</b>	<b>H1</b>	<b>H2</b>	<b>H3</b>
demonstrated	-0,3	-0,8	-0,3
ability	0,2	-0,5	-0,6
design	0,7885283763	-0,5938901483	-0,222729064
developmentally	-0,8	0	0,9
appropriate	-0,3	0,2	-0,8
lesson	-0,8	-0,3	0,2
activity	-0,6	0,2	-0,3
allow	0,2	-0,3	-0,8
integration	-0,3	0,2	-0,6
learning	-0,8	-0,3	0,2
style	-0,6	0,2	-0,3
highly	0,2	-0,3	-0,8
educate	-0,3	0,2	-0,6
differentiated	-0,8	-0,3	0,2
classroom	-0,6	0,2	-0,3

**Tabel 4.42 Pembaharuan bobot *input layer-hidden layer* (lanjutan)**

Term	H1	H2	H3
determined	0,2	-0,3	-0,8
maximize	-0,3	-0,5	0,2
educational	0,2	-0,3	-0,5
achievement	-0,8	0,2	-0,3
student	-0,5	-0,3	0,2
trained	-0,3	0,2	-0,8
developmental	0,2	-0,3	-0,5
reading	-0,8	0,2	-0,3
assessments	-0,5	-0,3	0,2

Pembaharuan bobot dihitung menggunakan *learning rate* dan gradien dari *hidden layer* ke *input layer* dilakukan seperti pada Persamaan Persamaan 4.83 sampai dengan Persamaan 4.90.

$$W_{t+1} = W_t - (0,1 \times dU) \quad (4.83)$$

$$W_{t+1} = \begin{bmatrix} 0,3 & \dots & 0,3 \\ -0,4 & \dots & 0 \\ -0,1 & \dots & -0,6 \end{bmatrix} - \left( 0,1 \times \begin{bmatrix} 0,0537 & \dots & 0,0467 \\ -0,0403 & \dots & -0,0350 \\ -0,0134 & \dots & -0,0117 \end{bmatrix} \right) \quad (4.84)$$

$$W_{t+1_{1,1}} = 0,3 - (0,1 \times 0,05372842428) = 0,2946271576 \quad (4.85)$$

$$W_{t+1_{1,24}} = 0,3 - (0,1 \times 0,04670924812) = 0,2953290752 \quad (4.86)$$

$$W_{t+1_{2,1}} = -0,4 - (0,1(-0,04029631821)) = -0,3959703682 \quad (4.87)$$

$$W_{t+1_{2,24}} = 0 - (0,1(-0,03503193609)) = 0,003503193609 \quad (4.88)$$

$$W_{t+1_{3,1}} = -0,1 - (0,1(-0,01343210607)) = -0,09865678939 \quad (4.89)$$

$$W_{t+1_{3,24}} = -0,6 - (0,1(-0,01167731203)) = -0,5988322688 \quad (4.90)$$

Hasil pembaharuan bobot *hidden layer* ke *output layer* tertera pada Tabel 4.43 hingga Tabel 4.46.

**Tabel 4.43 Pembaharuan bobot *input layer-output layer***

Neuron	demonstrated	ability	design	developmentally	appropriate	lesson
H1	0,294627	-0,16390	-0,20360	0,037385	-0,20319	-0,10319
H2	-0,39597	-0,32707	-0,5973	0,271961	-0,09760	-0,19760
H3	-0,09866	-0,80902	0,500900	0,190654	-0,39920	0,300799

**Tabel 4.44** Pembaharuan bobot *input layer-output layer*

<i>Neuron</i>	<i>activity</i>	<i>allow</i>	<i>integratio</i> <i>n</i>	<i>learning</i>	<i>style</i>	<i>highly</i>
H1	-0,10262	0,295235	-0,20237	-0,10319	-0,10262	0,295235
H2	0,301961	-0,19643	0,301775	-0,19760	0,301961	-0,19643
H3	-0,19935	-0,09881	-0,09941	0,3008	-0,19935	-0,09881

**Tabel 4.45** Pembaharuan bobot *input layer-output layer*

<i>Neuron</i>	<i>educate</i>	<i>differenti</i> <i>ated</i>	<i>classroo</i> <i>m</i>	<i>determin</i> <i>ed</i>	<i>maximize</i>	<i>educa</i> <i>tion</i>
H1	-0,20237	-0,10319	-0,10262	0,295235	-0,20237	-0,10319
H2	0,301775	-0,19760	0,301961	-0,19643	0,301775	-0,19760
H3	-0,09941	0,300799	-0,19935	-0,09881	-0,09941	0,300799

**Tabel 4.46** Pembaharuan bobot *input layer-output layer*

<i>Neuron</i>	<i>achievement</i>	<i>student</i>	<i>trained</i>	<i>developm</i> <i>ental</i>	<i>reading</i>	<i>assessme</i> <i>nts</i>
H1	-0,10262	0,295235	-0,20237	-0,40237	-0,10251	0,295329
H2	0,301961	-0,19643	0,301775	-0,09823	0,301885	0,003503
H3	-0,19935	-0,09881	-0,09941	0,300592	0,000628	-0,59883

#### 4.7.6 Perhitungan Manual Similaritas

Perhitungan manual similaritas dilakukan menggunakan vektor TF-IDF dan Word2Vec dari resume dengan ID 15265464 dan *section Experience*. Sedangkan untuk kualifikasi lowongan kerja menggunakan posisi *Teacher* dari PT Abadi Cahaya Edukasi. Isian data sampel yang dimaksud tertera pada Tabel 4.47.

**Tabel 4.47 Data sampel perhitungan manual similaritas**

Resume	Kualifikasi Lowongan Kerja
intervention specialist teacher math language arts identified student substantial academic difficulty evaluation use developmental reading assessments consultation staff member referred student developed differentiate lesson plan select appropriate instructional material reach individualized student goal developed implement creative lesson clear objective link common core incorporate differentiated instruction attended gain knowledge numerous service improved overall lexiles test score facilitated group lesson dependent student reading level determine run record assessment evaluated student growth progress monitoring formal informal assessment instructed student accordance schedule previously devise enhanced lesson use smart board technology computer assessed regular basis objective student set led basic skill class student time conducted small group individual classroom activity student base differentiated learning need nd grade replacement teacher implement positive behavior management use color system developed clear objective student parent lesson activity designed differentiated common core lesson plan activity meet need learner enhanced lesson use smartboard technology pads computer lab assessed student growth informal formal assessment developed lesson accordance student reading level determine quarterly running record testing maintained positive collaboration communication parent weekly newsletter weekly student progress update email conference attended service staff...	play vital role support lead teacher create nurturing stimulate learning environment young student assist lead teacher planning implement engage educational activity child age 36 year help maintain safe organised classroom environment provide individual attention support student needed collaborate teach team monitor record childrens progress participate staff meeting professional development opportunity maintain open communication parent caregiver ensure compliance relevant childcare regulation policy diploma degree early childhood education related field least 1 year experience work childcare educational setting strong interpersonal communication skill patience creativity genuine passion work young child ability work collaboratively part team knowledge child development ageappropriate teaching method proficiency bahasa indonesia english

#### 4.7.6.1 Perhitungan Manual TF-IDF Dengan *Improved Sqrt-Cosine Similarity*

Untuk metode representasi teks menggunakan TF-IDF, langkah pertama adalah mengambil nilai vektor  $x$  (resume) dan vektor  $y$  (kualifikasi lowongan kerja) dari hasil TF-IDF. Vektor resume dan kualifikasi lowongan kerja tertera pada Tabel 4.48.

**Tabel 4.48 Vektor TF-IDF perhitungan manual *Improved Sqrt-Cosine Similarity***

Resume	Kualifikasi Lowongan Kerja
[0 ... 0 ... 0,03772064 ... 0 ... 0 ... 0,07763420 ... 0 ... 0 ... 0,11012862 ... 0]	[0 ... 0 ... 0,06593613 ... 0 ... 0 ... 0,06276407 ... 0 ... 0 ... 0,09679995 ... 0]

Hitung penjumlahan dari akar perkalian elemen-elemen yang sesuai antara vektor  $x$  dan  $y$  sebagai numerator menggunakan formula seperti pada Persamaan 4.91.

$$\text{numerator} = \sum_{i=1}^m \sqrt{x_i y_i} \quad (4.91)$$

Demonstrasi perhitungan tertera mulai dari Persamaan 4.92 sampai dengan Persamaan 4.94.

$$\text{numerator} = \sqrt{x_1 y_1} + \sqrt{x_2 y_2} + \cdots + \sqrt{x_{526} y_{526}} + \cdots + \sqrt{x_{43331} y_{43331}} \quad (4.92)$$

$$\text{numerator} = \sqrt{0 \times 0} + \sqrt{0 \times 0} + \cdots + \sqrt{0,110 \times 0,063} + \cdots + \sqrt{0 \times 0} \quad (4.93)$$

$$\text{numerator} = 1,4044 \quad (4.94)$$

Hitung hasil kali dari akar penjumlahan semua elemen masing-masing vektor  $x$  (resume) dan vektor  $y$  (kualifikasi) sebagai denominator dengan formula seperti pada Persamaan 4.95.

$$\text{denominator} = \sqrt{(\sum_{i=1}^m x_i)} \sqrt{(\sum_{i=1}^m y_i)} \quad (4.95)$$

Demonstrasi perhitungan tertera mulai dari Persamaan 4.96 sampai dengan Persamaan 4.98.

$$\text{denominator} = \sqrt{(x_1 + x_2 + \cdots + x_{526} + \cdots + x_{43331})} \times \sqrt{(y_1 + y_2 + \cdots + y_{526} + \cdots + y_{43331})} \quad (4.96)$$

$$\text{denominator} = \sqrt{(0 + 0 + \cdots + 0,1101 + \cdots + 0)} \times \sqrt{(0 + 0 + \cdots + 0,0628 + \cdots + 0)} \quad (4.97)$$

$$\text{denominator} = 8,4315 \quad (4.98)$$

Skor similaritas *Improved Sqrt-Cosine* didapatkan dengan membagi numerator dengan denominator seperti formula pada Persamaan 4.99.

$$ISC = \frac{\text{numerator}}{\text{denominator}} \quad (4.99)$$

Hasil pembagian antara numerator dengan denominator yang sudah didapatkan tertera pada Persamaan 4.100.

$$ISC = \frac{1,4044}{8,4315} = 0,1666 \quad (4.100)$$

Hasil similaritas antara keseluruhan *section* resume ID 15265464 dengan kualifikasi lowongan kerja posisi *Teachers* tertera pada Tabel 4.49.

**Tabel 4.49 Hasil skor similaritas resume ID 15265464 perhitungan manual  
*Improved Sqrt-Cosine Similarity***

<i>Section</i>	Skor Similaritas
Education	0,1365
Experience	0,1666
Summary	0,1460

Pengujian pada penelitian ini menggunakan dua skenario, yakni “Tanpa Bobot” dan “Dengan Bobot”. Pada skenario “Tanpa Bobot”, skor similaritas untuk masing-masing *section* di resume ID 15265464 langsung dijumlah dan dirata-ratakan seperti pada Persamaan 4.101.

$$\text{Similarity} = \frac{0,1365+0,1666+0,1460}{3} = 0,1497 \quad (4.101)$$

Pada skenario “Dengan Bobot”, skor similaritas untuk masing-masing *section* di resume ID 15265464 dikalikan dengan persentase bobot yang diberikan oleh ahli, lalu ditotal dan dibagi dengan jumlah persentase *section* yang digunakan pada resume. Pada kategori industri “TEACHER”, diketahui rincian bobot untuk setiap *section* tertera pada Tabel 4.50.

**Tabel 4.50 Bobot *section* kategori industri “TEACHER”**

<i>Industry</i>	<i>Section</i>	<b>Bobot (%)</b>
TEACHER	Summary	5
TEACHER	Accomplishments/Awards	15
TEACHER	Skills/Qualifications	20
TEACHER	Education	20
TEACHER	Experience	20

**Tabel 4.50 Bobot section kategori industri "TEACHER" (lanjutan)**

Industry	Section	Bobot (%)
TEACHER	Organization	5
TEACHER	Projects	5
TEACHER	Certifications	10
TEACHER	Portfolio	10

Maka total skor similaritas untuk resume ID 15265464 dapat dihitung seperti pada Persamaan 4.102 sampai dengan Persamaan 4.103.

$$\text{Similarity} = \frac{(0,1365 \times 0,2) + (0,1666 \times 0,2) + (0,1460 \times 0,05)}{(0,2 + 0,2 + 0,05)} \quad (4.102)$$

$$\text{Similarity} = \frac{0,0273 + 0,03332 + 0,0073}{0,45} = 0,1509 \quad (4.103)$$

#### 4.7.6.2 Perhitungan Manual Word2Vec Dengan Cosine Similarity

Untuk metode representasi teks menggunakan Word2Vec, langkah pertama adalah mengambil nilai rata-rata dokumen untuk keseluruhan vektor  $x$  (resume) dan keseluruhan vektor  $y$  (kualifikasi lowongan kerja) dari hasil Word2Vec. Vektor resume dan kualifikasi lowongan kerja tertera pada Tabel 4.51.

**Tabel 4.51 Vektor Word2Vec perhitungan manual Cosine Similarity**

Resume	Kualifikasi
[0,071130395 0,12649915 -0,1145207 -0,08985851 ... 0,19086754]	[-0,04582439 0,0966028 -0,18032219 -0,11506447 ... 0,11572151]

Hitung penjumlahan dari perkalian elemen-elemen yang bersesuaian antara vektor  $x$  dan  $y$  sebagai numerator menggunakan formula seperti pada Persamaan 4.104.

$$\text{numerator} = \sum_{i=1}^m x_i y_i \quad (4.104)$$

Demonstrasi perhitungan tertera mulai dari Persamaan 4.105 sampai dengan Persamaan 4.107.

$$\text{numerator} = x_1 y_1 + x_2 y_2 + \dots + x_{100} y_{100} \quad (4.105)$$

$$\text{numerator} = (0,0711 \times (-0,0458)) + (0,1265 \times 0,0966) + \dots + (0,1909 \times 0,1157) \quad (4.106)$$

$$\text{numerator} = (-0,00323) + 0,01221 + \dots + 0,02208 = 1,9022 \quad (4.107)$$

Hitung hasil kali dari akar semua elemen masing-masing vektor  $x$  (resume) dan vektor  $y$  (kualifikasi) yang dikuadratkan sebagai denominator dengan formula seperti pada Persamaan 4.108.

$$denominator = \sqrt{(\sum_{i=1}^m x_i^2)} \sqrt{(\sum_{i=1}^m y_i^2)} \quad (4.108)$$

Demonstrasi perhitungan tertera mulai dari Persamaan 4.109 sampai dengan Persamaan 4.111.

$$denominator = \sqrt{(x_1^2 + x_2^2 + \dots + x_{100}^2)} \times \sqrt{(y_1^2 + y_2^2 + \dots + y_{100}^2)} \quad (4.109)$$

$$denominator = \sqrt{(0,0711^2 + 0,1265^2 + \dots + 0,1909^2)} \times \sqrt{((-0,0458)^2 + 0,0966^2 + \dots + 0,1157^2)} \quad (4.110)$$

$$denominator = 2,1667 \quad (4.111)$$

Skor *Cosine Similarity* didapatkan dengan membagi numerator dengan denominator seperti formula pada Persamaan 4.112.

$$CosSim = \frac{numerator}{denominator} \quad (4.112)$$

Hasil pembagian antara numerator dengan denominator yang sudah didapatkan tertera pada Persamaan 4.113.

$$CosSim = \frac{1,9022}{2,1667} = 0,8779 \quad (4.113)$$

Hasil similaritas antara keseluruhan *section* resume ID 15265464 dengan kualifikasi lowongan kerja posisi *Teachers* tertera pada Tabel 4.52.

**Tabel 4.52 Hasil skor similaritas resume ID 15265464 perhitungan manual  
*Cosine Similarity***

Section	Skor Similaritas
Education	0,7303
Experience	0,8779
Summary	0,9277

Pengujian pada penelitian ini menggunakan dua skenario, yakni “Tanpa Bobot” dan “Dengan Bobot”. Pada skenario “Tanpa Bobot”, skor similaritas untuk resume ID 15265464 bisa langsung dijumlah dan dirata-ratakan seperti pada Persamaan 4.114.

$$Similarity = \frac{0,7303+0,8779+0,9277}{3} = 0,8453 \quad (4.114)$$

Pada skenario “Dengan Bobot”, skor similaritas untuk masing-masing *section* di resume ID 15265464 dikalikan dengan persentase bobot yang diberikan oleh ahli. Pada kategori industri “TEACHER”, diketahui rincian bobot untuk setiap *section* tertera pada Tabel 4.50 sebelumnya.

Maka total skor similaritas untuk resume ID 15265464 dapat dihitung seperti pada Persamaan 4.115 sampai dengan Persamaan 4.116.

$$Similarity = \frac{(0,7303 \times 0,2) + (0,8779 \times 0,2) + (0,9277 \times 0,05)}{(0,2 + 0,2 + 0,05)} \quad (4.115)$$

$$Similarity = \frac{0,14606 + 0,17558 + 0,046385}{0,45} = 0,8178 \quad (4.116)$$

#### **4.7.6.3 Perhitungan Manual Word2Vec Dengan *Improved Sqrt-Cosine Similarity***

Pada perhitungan manual sebelumnya sudah dirincikan nilai rata-rata dokumen untuk keseluruhan vektor  $x$  (resume) dan keseluruhan vektor  $y$  (kualifikasi lowongan kerja) dari hasil Word2Vec yang tertera pada Tabel 4.51. Namun, untuk menghitung similaritasnya dengan *Improved Sqrt-Cosine*, perlu mengonversi nilai vektor-vektornya menjadi absolut karena akan menjadi bilangan imajiner jika nilai negatif diakarkan. Vektor resume dan kualifikasi lowongan kerja tertera pada Tabel 4.53.

**Tabel 4.53 Vektor Word2Vec perhitungan manual *Improved Sqrt-Cosine Similarity***

Resume	Kualifikasi
[0,071130395 0,12649915 0,1145207 0,08985851 ... 0,19086754]	[0,04582439 0,0966028 0,18032219 0,11506447 ... 0,11572151]

Hitung penjumlahan dari akar perkalian elemen-elemen yang sesuai antara vektor  $x$  dan  $y$  sebagai numerator menggunakan formula seperti pada Persamaan 4.117.

$$numerator = \sum_{i=1}^m \sqrt{x_i y_i} \quad (4.117)$$

Demonstrasi perhitungan tertera mulai dari Persamaan 4.118 sampai dengan Persamaan 4.120.

$$\text{numerator} = \sqrt{x_1 y_1} + \sqrt{x_2 y_2} + \cdots + \sqrt{x_{100} y_{100}} \quad (4.118)$$

$$\text{numerator} = \sqrt{0,0711 \times -0,0458} + \sqrt{0,1265 \times 0,0966} + \cdots + \sqrt{0,1909 \times 0,1157} \quad (4.119)$$

$$\text{numerator} = 11,2936 \quad (4.120)$$

Hitung hasil kali dari akar penjumlahan semua elemen masing-masing vektor  $x$  (resume) dan vektor  $y$  (kualifikasi) sebagai denominator dengan formula seperti pada Persamaan 4.121.

$$\text{denominator} = \sqrt{(\sum_{i=1}^m x_i)} \sqrt{(\sum_{i=1}^m y_i)} \quad (4.121)$$

Demonstrasi perhitungan tertera mulai dari Persamaan 4.122 sampai dengan Persamaan 4.124.

$$\text{denominator} = \sqrt{(x_1 + x_2 + \cdots + x_{100})} \times \sqrt{(y_1 + y_2 + \cdots + y_{100})} \quad (4.122)$$

$$\text{denominator} = \sqrt{(0,0711 + 0,1265 + \cdots + 0,1909)} \times \sqrt{(0,0458 + 0,0966 + \cdots + 0,1157)} \quad (4.123)$$

$$\text{denominator} = 11,861 \quad (4.124)$$

Skor similaritas *Improved Sqrt-Cosine* didapatkan dengan membagi numerator dengan denominator seperti formula pada Persamaan 4.125.

$$ISC = \frac{\text{numerator}}{\text{denominator}} \quad (4.125)$$

Hasil pembagian antara numerator dengan denominator yang sudah didapatkan tertera pada Persamaan 4.126.

$$ISC = \frac{11,2936}{11,861} = 0,9521 \quad (4.126)$$

Hasil similaritas antara keseluruhan *section* resume ID 15265464 dengan kualifikasi lowongan kerja posisi *Teachers* tertera pada Tabel 4.54.

**Tabel 4.54 Hasil skor similaritas resume ID 15265464 perhitungan manual  
*Improved Sqrt-Cosine Similarity***

Section	Skor Similaritas
Education	0,9065

**Tabel 4.54 Hasil skor similaritas resume ID 15265464 perhitungan manual**  
***Improved Sqrt-Cosine Similarity***

Section	Skor Similaritas
Experience	0,9521
Summary	0,9657

Pengujian pada penelitian ini menggunakan dua skenario, yakni “Tanpa Bobot” dan “Dengan Bobot”. Pada skenario “Tanpa Bobot”, skor similaritas untuk resume ID 15265464 bisa langsung dijumlah dan dirata-ratakan seperti pada Persamaan 4.127.

$$\text{Similarity} = \frac{0,9065+0,9521+0,9657}{3} = 0,9414 \quad (4.127)$$

Pada skenario “Dengan Bobot”, skor similaritas untuk masing-masing *section* di resume ID 15265464 dikalikan dengan persentase bobot yang diberikan oleh ahli. Pada kategori industri “TEACHER”, diketahui rincian bobot untuk setiap *section* tertera pada Tabel 4.50 sebelumnya.

Maka total skor similaritas untuk resume ID 15265464 dapat dihitung seperti pada Persamaan 4.128 sampai dengan Persamaan 4.129.

$$\text{Similarity} = \frac{(0,9065 \times 0,2) + (0,9521 \times 0,2) + (0,9657 \times 0,05)}{(0,2 + 0,2 + 0,05)} \quad (4.128)$$

$$\text{Similarity} = \frac{0,1813 + 0,19042 + 0,048285}{0,45} = 0,93334 \quad (4.129)$$

#### 4.7.7 Skenario Pengujian

Pengujian akan dilakukan dengan *human-level performance*, dihitung parameter setiap parameter penilaianya, mulai dari korelasi peringkat dengan *Spearman Rank Correlation Coefficient* (SRCC), persentase relevansi, dan persentase senioritas.

Setelah implementasi metode menghasilkan skor similaritas resume terhadap setiap kualifikasi lowongan kerja, diberikan pemeringkatan berdasarkan skor similaritas tertinggi, lalu diambil lima resume teratas. Kemudian, lima resume tersebut diberikan kepada seorang ahli di bidang rekrutmen untuk dinilai relevansinya secara manual. Ahli akan memberikan peringkat secara ulang sebagai *ground truth* dari peringkat satu sampai lima yang telah dihasilkan implementasi metode terhadap kualifikasi lowongan kerja tersebut. Selain itu, ahli akan mengevaluasi lima resume tersebut untuk memastikan relevansi dengan kualifikasi lowongan kerja yang tertera pada deskripsi lowongan kerja serta kesesuaian tingkat senioritas atau level posisi dengan deskripsi kualifikasi lowongan kerja.

Korelasi antara peringkat yang dihasilkan oleh implementasi metode dan peringkat dari ahli menunjukkan sejauh mana implementasi metode dapat meniru penilaian manusia. Semakin tinggi nilai korelasi (mendekati 1), semakin baik kinerja implementasi metode dalam menyelaraskan hasilnya dengan penilaian ahli. Untuk setiap implementasi metode, dilakukan perhitungan SRCC yang sama, maka untuk demonstrasi perhitungan SRCC akan menggunakan hasil pemeringkatan dari implementasi TF-IDF dengan *Improved Sqrt-Cosine Similarity* (ISC) pada kualifikasi lowongan kerja posisi “Teachers” skenario “Tanpa Bobot” yang tertera pada Tabel 4.55.

**Tabel 4.55 Peringkat 1-5 resume dengan skor similaritas terbesar untuk perhitungan manual SRCC**

Resume_ID	Peringkat	Peringkat Ahli
15850434	1	3
96547039	2	1
28772892	3	4
22056333	4	2
37220856	5	5

Setelah diketahui peringkat lima resume dengan skor similaritas terbesar hasil implementasi metode dan peringkat dari ahli, selanjutnya adalah menghitung selisih peringkat dan dikuadratkan yang didemonstrasikan pada Tabel 4.56.

**Tabel 4.56 Selisih peringkat 1-5 resume perhitungan manual SRCC**

Resume_ID	Peringkat	Peringkat Ahli	Selisih ( $d_i$ )	$d_i^2$
15850434	1	3	$1 - 3 = -2$	4
96547039	2	1	$2 - 1 = 1$	1
28772892	3	4	$3 - 4 = -1$	1
22056333	4	2	$4 - 2 = 2$	4
37220856	5	5	$5 - 5 = 0$	0

Pada Tabel 4.56, diketahui hasil kuadrat dari selisih perangkat. Jumlahkan hasil kuadrat tersebut untuk perhitungan SRCC sebagai numerator seperti pada Persamaan 4.130.

$$\sum d_i^2 = 4 + 1 + 1 + 4 + 0 = 10 \quad (4.130)$$

Gunakan rumus SRCC seperti Persamaan 4.131 untuk menghitung korelasi antara peringkat hasil implementasi dengan peringkat ahli.

$$SRCC = 1 - \frac{6 \sum d_i}{n(n^2-1)} \quad (4.131)$$

Demonstrasi perhitungan untuk peringkat lima resume dijabarkan mulai dari Persamaan 4.132 sampai dengan Persamaan 4.135.

$$SRCC = 1 - \frac{6 \times 10}{5(5^2-1)} \quad (4.132)$$

$$SRCC = 1 - \frac{60}{5(24)} \quad (4.133)$$

$$SRCC = \frac{120}{120} - \frac{60}{120} \quad (4.134)$$

$$SRCC = \frac{60}{120} = 0,5 \quad (4.135)$$

Hasil evaluasi relevansi dari ahli dihitung dalam bentuk persentase. Semakin besar persentase maka semakin baik suatu metode menghasilkan lima resume yang relevan dengan deskripsi lowongan kerja. Sama halnya dengan hasil evaluasi senioritas dari ahli, semakin besar persentase maka semakin baik suatu metode mengenali lima resume dengan level posisi yang sesuai. Untuk setiap implementasi metode, dilakukan perhitungan relevansi dan senioritas yang sama, maka untuk demonstrasi perhitungan relevansi dan senioritas akan menggunakan hasil pemeringkatan dari implementasi TF-IDF dengan *Improved Sqrt-Cosine Similarity* (ISC) pada kualifikasi lowongan kerja posisi “Teachers” skenario “Tanpa Bobot” yang tertera pada Tabel 4.57

**Tabel 4.57 Peringkat 1-5 resume dengan hasil evaluasi relevansi dan senioritas ahli terbesar untuk perhitungan manual**

Resume_ID	Relevance	Seniority
15850434	TRUE	TRUE
96547039	TRUE	TRUE
28772892	TRUE	TRUE
22056333	TRUE	TRUE
37220856	FALSE	TRUE

Untuk masing-masing parameter (relevansi dan senioritas), hitung pembagian antara jumlah nilai “TRUE” dengan jumlah total resume, lalu kalikan 100% seperti pada Persamaan 4.136 untuk relevansi dan 4.137 untuk senioritas.

$$P_{Relevance} = \left( \frac{N_r}{N} \right) \times 100\% \quad (4.136)$$

$$P_{Seniority} = \left( \frac{N_s}{N} \right) \times 100\% \quad (4.137)$$

Keterangan:

$P_{Relevance}$  = Persentase relevansi

$P_{Seniority}$  = Persentase senioritas

$N_r$  = Jumlah resume dengan relevansi bernilai TRUE

$N_s$  = Jumlah resume dengan senioritas bernilai TRUE

$N$  = Total jumlah resume

Demonstrasi perhitungan untuk relevansi lima resume tertera pada Persamaan 4.138.

$$P_{Relevance} = \left( \frac{4}{5} \right) \times 100\% = 0,8 \times 100\% = 80\% \quad (4.138)$$

Demonstrasi perhitungan untuk senioritas lima resume tertera pada Persamaan 4.139.

$$P_{Seniority} = \left( \frac{5}{5} \right) \times 100\% = 1 \times 100\% = 100\% \quad (4.139)$$

## BAB 5 IMPLEMENTASI

Bab implementasi berisi implementasi dari metode kalkulasi similaritas teks pada resume pelamar dengan kualifikasi instansi yang digunakan pada penelitian ini, seperti metode perhitungan similaritas *Improved Sqrt-Cosine* (ISC) dan *Cosine Similarity* (CosSim), serta metode representasi teks TF-IDF dan Word2Vec.

### 5.1 Implementasi Kode Program *Import Libraries* dan *Load Dataset*

Dalam kode program ini, dilakukan pemuatan *dataset* resume yang akan dikelola dan *dataset* kualifikasi lowongan kerja yang dikumpulkan dari <https://id.jobstreet.com/> dengan 24 posisi kualifikasi lowongan kerja. Implementasi kode program tertera pada Kode Program 5.1.

#### Kode Program 5.1 Implementasi kode program *import libraries* dan *load dataset*

```
1 import pandas as pd
2 from bs4 import BeautifulSoup
3 import re
4 from gensim.models import Word2Vec
5 import numpy as np
6 from sklearn.metrics.pairwise import cosine_similarity
7 from sklearn.feature_extraction.text import TfidfVectorizer
8 import string
9 from tqdm import tqdm
10 import time
11
12 import nltk
13 nltk.download('wordnet')
14 nltk.download('omw-1.4')
15 nltk.download('punkt')
16 nltk.download('averaged_perceptron_tagger')
17 nltk.download('averaged_perceptron_tagger_eng')
18
19 from nltk.corpus import wordnet
20 from nltk.stem import WordNetLemmatizer
21 from nltk.tokenize import word_tokenize
22 from nltk.corpus import stopwords
23 from nltk import pos_tag
24
25 # Load dataset resume
26 resume_df =
27 pd.read_csv(r'C:\...\archive2024\Resume\Resume.csv')
28 resume_df
29
30 # Load dataset kualifikasi lowongan kerja
31 vacancy_df =
32 pd.read_csv(r'C:\...\archive2024\kualifikasi_loker.csv')
vacancy_df
```

### Kode Program 5.1 Implementasi kode program import libraries dan load dataset (lanjutan)

```
33     # Load dataset bobot section
34     section_df =
35     pd.read_csv(r'C:\...\archive2024\bobot_section.csv')
            section_df
```

Penjelasan dari Kode Program 5.1 mengenai implementasi *import libraries* dan *load dataset*, yaitu:

1. Baris 1-10 merupakan proses *import library* yang diperlukan untuk:
  - a. pandas alias pd untuk manipulasi data dan analisis data,
  - b. BeautifulSoup dari *library* bs4 untuk ekstraksi teks dari HTML,
  - c. re untuk ekspresi reguler atau *regular expression* (REGEX),
  - d. gensim.models import Word2Vec untuk menggunakan *library* Gensim dalam implementasi representasi teks menggunakan pendekatan Word2Vec,
  - e. numpy alias np untuk operasi numerik seperti perhitungan matematis,
  - f. sklearn.metrics.pairwise import cosine\_similarity untuk menghitung kesamaan antar vektor teks menggunakan *Cosine Similarity* dari *library* Scikit-learn,
  - g. sklearn.feature\_extraction.text import TfidfVectorizer untuk menggunakan *library* Scikit-learn dalam implementasi representasi teks menggunakan pendekatan TF-IDF,
  - h. import string untuk menyediakan daftar karakter tanda baca untuk *preprocessing* teks,
  - i. from tqdm import tqdm untuk menampilkan *progress bar* guna memantau *progress* dari suatu proses,
  - j. import time untuk mengukur waktu eksekusi.
2. Baris 12-17 merupakan proses *import library* dari nltk (Natural Language Toolkit) yang digunakan untuk pemrosesan teks, termasuk unduhan *resource* yang diperlukan untuk tokenisasi, lematisasi, dan *Part of Speech (POS) tagging*, seperti:
  - a. wordnet merupakan basis data leksikal Bahasa Inggris,
  - b. omw-1.4 merupakan *Open Multilingual Wordnet* versi 1.4 untuk mendukung lematisasi dalam berbagai bahasa,
  - c. Punkt untuk memecah teks atau tokenisasi teks
  - d. averaged\_perceptron\_tagger dan averaged\_perceptron\_tagger\_eng untuk POS *tagging* yang memberi label jenis kata seperti kata benda (*noun*), kata kerja (*verb*), kata sifat (*adjective*), dan kata keterangan (*adverb*).

3. Baris 19-23 merupakan *import* tambahan dari *library nltk*, yaitu:
  - a. wordnet dari `nltk.corpus` untuk mengakses basis data leksikal WordNet,
  - b. `WordNetLemmatizer` dari `nltk.stem` untuk melakukan mengubah kata ke bentuk dasar atau lematisasi,
  - c. `word_tokenize` dari `nltk.tokenize` untuk memecah teks menjadi token,
  - d. `stopwords` dari `nltk.corpus` menyediakan daftar kata-kata umum yang tidak bermakna dalam analisis teks untuk dihapus,
  - e. `pos_tag` untuk memberikan label POS *tagging* pada setiap kata dalam teks.
4. Baris 25-26 merupakan proses memuat *dataset* dari *file* dengan format .csv bernama `Resume.csv` menggunakan `pandas.read_csv()` yang berisi informasi resume kandidat dan mengubahnya menjadi DataFrame.
5. Baris 27 merupakan sintaksis untuk menampilkan DataFrame `resume_df`.
6. Baris 29-30 merupakan proses memuat *dataset* dari *file* dengan format .csv bernama `kualifikasi_loker.csv` menggunakan `pandas.read_csv()` dan mengubahnya menjadi DataFrame yang berisi informasi lowongan pekerjaan, termasuk nama posisi, nama perusahaan, dan deskripsi kualifikasinya.
7. Baris 31 merupakan sintaksis untuk menampilkan DataFrame `vacancy_df`.
8. Baris 33-34 merupakan proses memuat *dataset* dari *file* dengan format .csv bernama `bobot_section.csv` menggunakan `pandas.read_csv()` dan mengubahnya menjadi DataFrame yang berisi bobot pemberian ahli untuk setiap *section* yang ada di resume.

## 5.2 Implementasi Kode Program *Preprocessing* Resume

Dalam kode program ini, dilakukan untuk melakukan pra-pemrosesan *dataset* Resume. Diawali dengan mengekstrak setiap *section* dari resume yang berformat HTML dan disimpan dalam bentuk DataFrame dengan tambahan informasi mengenai bagian (*section*) dan isi teks dari masing-masing *section*. Implementasi kode program ekstraksi *section* tertera pada Kode Program 5.2.

### Kode Program 5.2 Implementasi kode program *preprocessing* resume bagian ekstraksi *section*

```
1 # Melihat informasi kolom dan tipe data
2 resume_df.info()
3
4 # Cek missing values
5 resume_df.isnull().sum()
6
7 # Menghapus kolom yang tidak digunakan
8 resume_df_1 = resume_df.drop(columns=["Category"])
9
10 # List untuk menyimpan hasil sementara per section dengan semua
11 # data dari df
12 data = []
13
14 # Loop untuk memproses setiap resume
15 for index, row in resume_df.iterrows():
16     # Ambil ID dan data lainnya dari DataFrame yang ada
17     resume_data = row.to_dict() # Mengambil semua data di
18     # baris ini
19
20     # Ambil resume_str langsung dari kolom 'Resume_str'
21     resume_str = row['Resume_str']
22
23     # Menemukan semua <div> dengan class "sectiontitle" di
24     # 'Resume_html'
25     soup = BeautifulSoup(row['Resume_html'], "html.parser")
26     section_divs = soup.find_all("div", class_="sectiontitle")
27
28     # Menyimpan teks dari setiap section
29     sections = [div.get_text(strip=True) for div in
30     section_divs]
31
32     # Menemukan posisi setiap section dalam resume_str
33     for i, section in enumerate(sections):
34         # Copy data resume agar setiap section mendapatkan data
35         # asli resume
36         section_data = resume_data.copy()
37
38         # Cari posisi awal section
39         start_index = resume_str.find(section)
40
41         # Tentukan posisi akhir section
42         if i + 1 < len(sections):
43             end_index = resume_str.find(sections[i + 1],
44             start_index)
```

## Kode Program 5.2 Implementasi kode program *preprocessing* resume bagian ekstraksi *section* (lanjutan)

```
39         else:
40             end_index = len(resume_str) # Jika ini adalah
41             section terakhir
42
43             # Ambil teks dari section tersebut dan hapus nama
44             section jika ada di awal teks
45             section_text =
46             resume_str[start_index:end_index].strip()
47             if section_text.startswith(section):
48                 section_text = section_text[len(section):].strip()
49
50             # Tambahkan kolom Section dan Text
51             section_data["Section"] = section
52             section_data["Text"] = section_text
53
54             # Tambahkan hasil ke expanded_data
55             data.append(section_data)
56
57             # Mengonversi list menjadi DataFrame
58             resume_df_1 = pd.DataFrame(data)
```

Penjelasan dari Kode Program 5.2 mengenai implementasi kode program ekstraksi *section*, yaitu:

1. Baris 1-2 merupakan proses untuk menampilkan informasi mengenai *dataset* resume, termasuk jumlah kolom, nama kolom, tipe data, dan jumlah *non-null* menggunakan `df.info()`.
2. Baris 4-5 merupakan proses untuk memeriksa nilai yang hilang (*missing values*) dalam *dataset* resume menggunakan `df.isnull().sum()`.
3. Baris 7-8 merupakan proses untuk menghapus kolom *Category* dari *dataset* resume menggunakan `df.drop(columns=["Category"])` dan menyimpan hasilnya ke DataFrame baru sebagai `resume_df_1`.
4. Baris 10-11 merupakan proses untuk membuat *list* kosong bernama `data` untuk menyimpan hasil pemrosesan ekstraksi *section* resume.
5. Baris 13-14 merupakan proses untuk memulai iterasi melalui setiap baris *dataset* resume menggunakan `df.iterrows()`.
6. Baris 15-16 merupakan proses untuk mengubah data baris menjadi *dictionary* menggunakan `row.to_dict()` dan menyimpannya ke variabel `resume_data`.
7. Baris 18-19 merupakan proses untuk mengambil teks resume dari kolom `Resume_str` menggunakan `row['Resume_str']` dan menyimpannya ke variabel `resume_str`.
8. Baris 21-23 merupakan proses untuk mem-parsing kolom `Resume_html` menggunakan `BeautifulSoup(row['Resume_html'], "html.parser")` dan menemukan semua elemen `<div>` dengan kelas `sectiontitle` menggunakan `soup.find_all()`.

9. Baris 25-26 merupakan proses untuk mengekstrak teks dari setiap elemen <div> menggunakan div.get\_text(strip=True) dan menyimpannya dalam *list* sections.
10. Baris 28-29 merupakan proses untuk memulai iterasi melalui setiap section dalam *list* sections menggunakan enumerate(sections).
11. Baris 30-31 merupakan proses untuk membuat salinan *dictionary* resume\_data menggunakan resume\_data.copy() dan menyimpannya ke variabel section\_data.
12. Baris 33-34 merupakan proses untuk mencari posisi awal nama section dalam teks resume\_str menggunakan resume\_str.find(section).
13. Baris 36-40 merupakan proses untuk menentukan posisi akhir section dengan memeriksa apakah ada section berikutnya menggunakan resume\_str.find(sections[i + 1], start\_index) atau menggunakan panjang resume\_str jika section terakhir.
14. Baris 42-45 merupakan proses untuk mengambil teks section dari resume\_str menggunakan *slicing* [start\_index:end\_index], menghapus spasi berlebih dengan strip(), dan menghapus nama section jika ada yang terdeteksi di awal kalimat pada teks menggunakan section\_text[len(section):].strip().
15. Baris 47-49 merupakan proses untuk menambahkan key Section dan Text ke *dictionary* section\_data dengan value nama section dan teks section yang telah diekstrak.
16. Baris 51-52 merupakan proses untuk menambahkan *dictionary* section\_data ke *list* data menggunakan data.append().
17. Baris 54-55 merupakan proses untuk mengonversi *list* data menjadi DataFrame baru menggunakan pd.DataFrame(data) dan menyimpannya sebagai resume\_df\_1.

Selanjutnya, isian resume dilakukan langkah-langkah pra-pemrosesan teks seperti yang sudah dijelaskan pada diagram alur di bab Perancangan. Implementasi kode program *preprocessing* isian resume tertera pada Kode Program 5.3.

### Kode Program 5.3 Implementasi kode program *preprocessing* resume bagian *preprocessing* isian resume

```

1      # Inisialisasi lemmatizer
2      lemmatizer = WordNetLemmatizer()
3
4      # Daftar stop words
5      stop_words = set(stopwords.words('english'))
6
7      # Fungsi untuk mendapatkan tipe kata untuk lemmatization
8      def get_wordnet_pos(tag):
9          if tag.startswith('J'):
10              return wordnet.ADJ
11          elif tag.startswith('V'):
12              return wordnet.VERB
13          elif tag.startswith('N'):
14              return wordnet.NOUN
15          elif tag.startswith('R'):
16              return wordnet.ADV
17          else:
18              return wordnet.NOUN
19
20      def preprocess(text):
21          # Hapus email dan nomor telepon
22          email_pattern = r'\b[A-Za-z0-9._%+-]+@[A-Za-z0-9.-]+\.[A-
Za-z]{2,}\b'
23          phone_pattern = r'(\b(?:\d{1,3}[-\.\s]?)?(?:\d{2,4})?[-\.\s]?)?\d{2,4}[-\.\s]?\d{2,4}[-\.\s]?\d{2,4}\b'
24
25          text = re.sub(email_pattern, '', text)
26          text = re.sub(phone_pattern, '', text)
27
28          # Hapus berbagai tipe tanda minus
29          text = re.sub(r'[\u2010-
\u2015\u2212\uFF0D\uFF0E\uFE63\u002D]', ' ', text)
30
31          # Regex untuk menghapus bulan (termasuk singkatan) &
32          # present/current
33          bulan_pattern =
r"\b(?:january|jan|february|feb|march|mar|april|apr|may|june|ju
n|july|jul|august|aug|september|sep|october|oct|november|nov|de
cember|dec)\b"
34          present_pattern = r"\b(?:present|current)\b"
35
36          text = re.sub(bulan_pattern, '', text, flags=re.IGNORECASE)
# Hapus bulan dan singkatan
37          text = re.sub(present_pattern, '', text,
flags=re.IGNORECASE) # Hapus "present/current"
38
39          # Regex untuk menangani berbagai format tanggal & rentang

```

### Kode Program 5.3 Implementasi kode program *preprocessing* resume bagian *preprocessing* isian resume (lanjutan)

```
39     date_pattern = r"""
40         \b(
41            (?:\d{1,2}/(?:\d{4}|Current)) # Format "01/2024" atau
42             "01/ Current"
43             |(?:\d{4})                                # Tahun "2023"
44             |(?:\s?(?:-|to|-|-)\s?(?:\d{4}|Current|Present))? #
45             Rentang waktu "2022 - 2023" atau "2022 - Present"
46             )\b
47         """
48
49     text = re.sub(date_pattern, '', text, flags=re.IGNORECASE |
50                   re.VERBOSE)
51
52     # Hapus placeholder seperti "Company Name" dan "State"
53     text = re.sub(r'\b(?:Company Name|State|City)\b', '', text,
54                   flags=re.IGNORECASE)
55
56     # Hapus tanda baca
57     text = re.sub(r'[^a-zA-Z\s]', '', text)
58
59     # Hapus angka
60     text = re.sub(r'\d+', '', text)
61
62     # Hapus spasi berlebihan setelah penghapusan
63     text = re.sub(r'\s+', ' ', text).strip()
64
65     tokens = word_tokenize(text)
66     tokens_pos = pos_tag(tokens)
67
68     # Lemmatization dan hapus stop words
69     lemmatized_text = []
70     for token, pos in tokens_pos:
71         if token.lower() not in stop_words: # Menghapus stop
72             words
73                 wordnet_pos = get_wordnet_pos(pos) or wordnet.NOUN
74                 lemmatized_text.append(lemmatizer.lemmatize(token,
pos=wordnet_pos))
75
76             return ' '.join(lemmatized_text)
77
78     # Implementasikan preprocessing ke dataset
79     resume_df_1['Text'] = resume_df_1['Text'].apply(preprocess)
80     resume_df_1['Text'] = resume_df_1['Text'].apply(preprocess)
```

Penjelasan dari Kode Program 5.3 mengenai implementasi kode program ekstraksi *section*, yaitu:

1. Baris 1-2 merupakan proses untuk menginisialisasi object WordNetLemmatizer dari NLTK menggunakan WordNetLemmatizer() dan menyimpannya ke variabel lemmatizer untuk melakukan lemmatisasi kata.
2. Baris 4-5 merupakan proses untuk membuat *set stop words* dalam Bahasa Inggris menggunakan stopwords.words('english') dan menyimpannya ke variabel stop\_words.

3. Baris 7-18 merupakan proses untuk mendefinisikan fungsi `get_wordnet_pos` yang mengonversi *tag part-of-speech* (POS) ke tipe kata WordNet seperti `wordnet.ADJ`, `wordnet.VERB`, `wordnet.NOUN`, `wordnet.ADV` berdasarkan awalan *tag*, dengan `default wordnet.NOUN` jika tidak sesuai.
4. Baris 20 merupakan pendefinisian fungsi `preprocess` yang menerima parameter `text` untuk mengeksekusi proses *preprocessing* teks.
5. Baris 21-23 merupakan proses untuk mendefinisikan pola *regex* untuk *email* menggunakan `email_pattern` dan nomor telepon menggunakan `phone_pattern` untuk dihapus dari teks.
6. Baris 25-26 merupakan proses untuk menghapus *email* dan nomor telepon dari teks menggunakan `re.sub()` dengan pola *regex* yang telah didefinisikan sebelumnya.
7. Baris 28-29 merupakan proses untuk menghapus berbagai jenis tanda minus dari teks menggunakan `re.sub()` dan menggantinya dengan spasi.
8. Baris 31-33 merupakan proses untuk mendefinisikan pola *regex* untuk menghapus nama bulan menggunakan `bulan_pattern` dan kata "*present*" atau "*current*" menggunakan `present_pattern`.
9. Baris 35-36 merupakan proses untuk menghapus nama bulan dan kata "*present*" atau "*current*" dari teks menggunakan `re.sub()` dengan pola *regex*, mengabaikan huruf besar maupun kecil dengan `flags=re.IGNORECASE`.
10. Baris 38-46 merupakan proses untuk mendefinisikan pola *regex* untuk berbagai format tanggal menggunakan `date_pattern` dan menghapusnya dari teks menggunakan `re.sub()` dengan flag `re.IGNORECASE` dan `re.VERBOSE`.
11. Baris 48-49 merupakan proses untuk menghapus kata-kata *placeholder* seperti "*Company Name*", "*State*", dan "*City*" dari teks menggunakan `re.sub()` dengan `flags=re.IGNORECASE`.
12. Baris 51-52 merupakan proses untuk menghapus semua tanda baca dan karakter non-huruf.
13. Baris 54-55 merupakan proses untuk menghapus semua angka dari teks.
14. Baris 57-58 merupakan proses untuk mengganti spasi yang berlebihan menjadi hanya satu spasi dan menggunakan `strip()` untuk menghapus spasi di awal dan akhir *string*.
15. Baris 60 merupakan proses untuk memecah teks menjadi daftar kata (*token*) menggunakan `word_tokenize()` dari NLTK.
16. Baris 61 merupakan proses untuk memberikan *part-of-speech* (POS) *tag* pada setiap token menggunakan `pos_tag()` dari NLTK.
17. Baris 63-68 merupakan proses untuk membuat *list* kosong `lemmatized_text`, mengiterasi token dan *tag* POS-nya, menghapus *stop words* jika token tidak ada di `stop_words`, mengonversi *tag* POS ke format WordNet menggunakan `get_wordnet_pos()`, dan melakukan lematisasi menggunakan `lemmatizer.lemmatize()`.

18. Baris 70 merupakan proses untuk menggabungkan token yang telah di-lematiasi menjadi satu *string* dengan spasi sebagai pemisah menggunakan ' '.join().
19. Baris 72-74 merupakan proses untuk menerapkan fungsi preprocess ke kolom Text pada DataFrame resume\_df\_1 menggunakan df['Text'].apply(preprocess). Proses ini dilakukan dua kali karena pada iterasi pertama ada beberapa nama bulan yang belum sepenuhnya terhapus, lalu pada iterasi kedua menghasilkan teks bersih yang sudah sesuai.

Selanjutnya adalah menstandarisasi kolom Section ke huruf kecil, menyeragamkan nama *section* menggunakan keyword\_mapping, menggabungkan teks yang diketahui *section*-nya lebih dari satu berdasarkan per ID resume-nya, melakukan pemetaan untuk *section* yang tidak valid berdasarkan pola teks, dan menghapus baris dengan teks kosong. Implementasi kode program *preprocessing* isian resume per *section* tertera pada Kode Program 5.4.

**Kode Program 5.4 Implementasi kode program *preprocessing* resume bagian *preprocessing* isian resume per *section***

```

1  resume_df_1["Section"] = resume_df_1["Section"].str.lower()
2
3  # Menampilkan nilai unik dari kolom 'Section'
4  unique_sections = resume_df_1['Section'].unique()
5
6  # Menampilkan hasil
7  print("Unique Sections in the Resume:")
8  for section in unique_sections:
9      print(section)
10
11 # Mapping kata kunci ke kategori yang diinginkan
12 keyword_mapping = {
13     "experience": "Experience",
14     "skill": "Skills/Qualifications",
15     "award": "Accomplishments/Awards",
16     "project": "Projects",
17     "education": "Education",
18     "certification": "Certifications",
19     "portfolio": "Portfolio",
20     "organization": "Organization",
21     "volunteer": "Organization",
22     "accomplishment": "Accomplishments/Awards",
23     "achievement": "Accomplishments/Awards",
24     "summary": "Summary",
25     "overview": "Summary",
26     "course": "Education",
27     "academ": "Education",
28     "work": "Experience",
29     "profile": "Summary",
30     "strength": "Skills/Qualifications",

```

**Kode Program 5.4 Implementasi kode program *preprocessing* resume bagian *preprocessing* isian resume per section (lanjutan)**

31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47 48 49 50 51 52 53 54 55 56 57 58 59 60 61 62 63 64 65 66 67 68 69 70 71 72 73 74 75 76 77 78 79 80 81 82 83 84 85 86	<pre>           "competencies": "Skills/Qualifications",           "competencies": "Skills/Qualifications",           "quali": "Skills/Qualifications",           "honor": "Certifications",           "honour": "Certifications",           "affiliation": "Certifications",           "affiliation": "Certifications",           "language": "Skills/Qualifications",           "community": "Organization",           "about": "Summary",           "training": "Certifications",           "scholarship": "Education",           "license": "Certifications",           "highlight": "Skills/Qualifications",           "expertise": "Skills/Qualifications",           "focus": "Summary",           "background": "Summary",           "interest": "Skills/Qualifications",           "military": "Experience",           "presentation": "Skills/Qualifications",           "objective": "Summary",           "reference": "Skills/Qualifications",           "referance": "Skills/Qualifications",           "proficien": "Skills/Qualifications",           "dissertation": "Projects",           "publications": "Skills/Qualifications",           "associat": "Skills/Qualifications",           "professional": "Experience",           "leadership": "Organization",           "curricular": "Organization",           "credential": "Skills/Qualifications",           "information": "Others",           "societ": "Organization",           "research": "Skills/Qualifications",           "employment": "Experience",           "adjunct": "Skills/Qualifications",           "personal": "Others",           "characteristic": "Others",           "goal": "Summary",           "apply": "Summary",           "role": "Experience",           "general": "Others",           "link": "Portfolio",           "snap shot": "Experience",           "tool": "Skills/Qualifications",           "hobb": "Others",           "activit": "Organization",           "client": "Experience",           "success": "Accomplishments/Awards",           "computer": "Skills/Qualifications",           "technical": "Skills/Qualifications",           "acumen": "Skills/Qualifications",           "development": "Skills/Qualifications",           "knowledge": "Skills/Qualifications",           "membership": "Skills/Qualifications",           "speak": "Accomplishments/Awards",         </pre>
--	---

### Kode Program 5.4 Implementasi kode program *preprocessing* resume bagian *preprocessing* isian resume per *section* (lanjutan)

```
87     "participat": "Projects",
88     "vocation": "Experience",
89     "clearance": "Skills/Qualifications",
90     "attribute": "Skills/Qualifications",
91     "exhibit": "Projects",
92 }
93
94 # Ubah nilai 'Section' berdasarkan kata kunci
95 for keyword, category in keyword_mapping.items():
96     resume_df_1.loc[resume_df_1['Section'].str.contains(keyword,
97     case=False, na=False), 'Section'] = category
98
99 # Gabungkan teks dari section terkait jika ada duplikasi dalam
100 # satu resume
101 resume_df_1 = resume_df_1.groupby(['ID', 'Section'],
102 as_index=False).agg({
103     'Text': ' '.join,
104     'Resume_str': 'first',
105     'Resume_html': 'first',
106 })
107
108 # Daftar section yang tidak boleh diubah
109 allowed_sections = [
110     "Summary", "Accomplishments/Awards",
111     "Skills/Qualifications",
112     "Education", "Experience", "Organization", "Projects",
113     "Certifications", "Portfolio", "Others"
114 ]
115
116 # Pola regex untuk kata-kata yang harus masuk ke
117 # Summary
118 summary_pattern = r'\b(?:Summary|I
am|I\m|experience|professional)\b'
119
120 # Ubah ke "Summary" jika ada salah satu kata dalam
121 # summary_pattern
122 resume_df_1.loc[
123     (~resume_df_1["Section"].isin(allowed_sections)) &
124     (resume_df_1["Text"].str.contains(summary_pattern,
125     case=False, na=False, regex=True)),
126     "Section"
127 ] = "Summary"
128
129 # Ubah ke "Portfolio" jika ada kata "LinkedIn" dalam
130 # "Text"
131 resume_df_1.loc[
132     (~resume_df_1["Section"].isin(allowed_sections)) &
133     (resume_df_1["Text"].str.contains(r'\bLinkedIn\b',
134     case=False, na=False, regex=True)),
135     "Section"
136 ] = "Portfolio"
```

#### Kode Program 5.4 Implementasi kode program *preprocessing* resume bagian *preprocessing* isian resume per *section* (lanjutan)

```
129     # Jika section tidak termasuk valid_sections dan kolom "Text"  
130     kosong, hapus baris tersebut  
131     resume_df_1 =  
132     resume_df_1[~((~resume_df_1['Section'].isin(allowed_sections))  
133     & (resume_df_1['Text'].isna()))]  
134     # Jika section tidak termasuk valid_sections tetapi ada isian  
135     di "Text", ubah menjadi "Others"  
136     resume_df_1.loc[~resume_df_1['Section'].isin(allowed_sections),  
137     'Section'] = 'Others'  
138     # Hapus baris dengan Text yang NaN, kosong, atau hanya spasi  
139     untuk semua Section  
140     resume_df_1 = resume_df_1[~(resume_df_1['Text'].isna() |  
141     (resume_df_1['Text'].str.strip() == ''))]  
142     resume_df_1["Text"] = resume_df_1["Text"].str.lower()  
143     resume_df_1 = resume_df_1.drop(columns=["Resume_str"])  
144     resume_df_1 = resume_df_1.drop(columns=["Resume_html"])
```

Penjelasan dari Kode Program 5.4 mengenai implementasi kode program ekstraksi *section*, yaitu:

1. Baris 1 merupakan proses untuk mengubah semua nilai di kolom Section pada DataFrame resume\_df\_1 menjadi huruf kecil menggunakan str.lower() guna lebih mudah dalam menyeragamkan nama-nama *section*.
2. Baris 3-9 merupakan proses untuk mengambil nilai unik dari kolom Section pada DataFrame resume\_df\_1 menggunakan unique() dan menyimpannya ke variabel unique\_sections guna memahami ada apa saja nama-nama *section* yang digunakan seluruh resume di dataset.
3. Baris 11-92 merupakan proses untuk mendefinisikan *dictionary keyword\_mapping* yang berisi pasangan *key* dan *value*, di mana *key* adalah nama *section* yang telah diketahui dari pemeriksaan nilai unik kolom Section sebelumnya, lalu *value* adalah nama *section* yang diinginkan untuk mengelompokkan nama-nama *section* resume.
4. Baris 94-96 merupakan proses untuk mengiterasi setiap pasangan *key* dan *value* di *keyword\_mapping* menggunakan items(), lalu mengubah nilai kolom Section yang mengandung kata kunci seperti di *key* menjadi *value* seperti *keyword\_mapping* dan mengabaikan huruf besar maupun kecil.
5. Baris 98-103 merupakan proses untuk mengelompokkan isi DataFrame resume\_df\_1 berdasarkan kolom ID dan Section menggunakan groupby(), lalu menggabungkan teks untuk kolom Text menggunakan ' '.join jika ada nilai di kolom Section yang sama.

6. Baris 105-110 merupakan proses untuk mendefinisikan *list allowed\_sections* yang berisi daftar nama-nama *section* yang diizinkan untuk tetap ada dalam DataFrame.
7. Baris 112-113 merupakan proses untuk mendefinisikan pola *regex summary\_pattern* yang mencakup kata-kata seperti "Summary", "I am", "I'm", "years", "experience", atau "professional" untuk mengidentifikasi teks yang relevan dengan nama *section* "Summary".
8. Baris 115-120 merupakan proses untuk mengubah nilai kolom Section menjadi "Summary" jika *section* tidak ada di *allowed\_sections* dan kolom Text mengandung kata-kata dari *summary\_pattern*.
9. Baris 122-127 merupakan proses untuk mengubah nilai kolom Section menjadi "Portfolio" jika *section* tidak ada di *allowed\_sections* dan kolom Text mengandung kata "LinkedIn".
10. Baris 129-130 merupakan proses untuk menghapus baris dari *resume\_df\_1* jika kolom Section tidak ada di *allowed\_sections* dan kolom Text kosong.
11. Baris 132-133 merupakan proses untuk mengubah nilai kolom Section menjadi "Others" jika *section* tidak ada di *allowed\_sections* tetapi kolom Text tidak kosong.
12. Baris 135-136 merupakan proses untuk menghapus baris dari *resume\_df\_1* jika kolom Text kosong atau jika hanya berisi spasi.
13. Baris 138 merupakan proses untuk mengubah semua nilai di kolom Text pada DataFrame *resume\_df\_1* menjadi huruf kecil atau *lower casing*.
14. Baris 139 merupakan proses untuk menghapus kolom *Resume\_str* dari DataFrame *resume\_df\_1* karena sudah tidak digunakan.
15. Baris 140 merupakan proses untuk menghapus kolom *Resume\_html* dari DataFrame *resume\_df\_1* karena sudah tidak digunakan.

### **5.3 Implementasi Kode Program *Preprocessing* Kualifikasi Lowongan Kerja**

Dalam kode program ini, dilakukan untuk membersihkan teks pada kolom *Description* yang merupakan deskripsi lowongan kerja dengan mengubah ke huruf kecil, menghapus tanda baca dan spasi berlebih, melakukan tokenisasi, POS tagging, menghapus *stop words*, dan lematisasi. Implementasi kode program tertera pada Kode Program 5.5.

#### **Kode Program 5.5 Implementasi kode program *preprocessing* kualifikasi lowongan kerja**

```

1 def preprocess_vacancy(text):
2     text = text.lower() # Ubah ke huruf kecil
3     text = re.sub(r'\d+', '', text) # Hapus angka
4     text = text.translate(str.maketrans('', '',
5         string.punctuation)) # Hapus tanda baca
6     text = ' '.join(text.split()) # Hapus spasi berlebih

```

### Kode Program 5.5 Implementasi kode program *preprocessing* kualifikasi lowongan kerja (lanjutan)

```
7      # Tokenisasi
8      tokens = word_tokenize(text)
9
10     # POS tagging
11     tagged_tokens = pos_tag(tokens)
12
13     # Hapus stop words & Lemmatization
14     processed_tokens = [lemmatizer.lemmatize(word,
15                                         get_wordnet_pos(tag))
16                                         for word, tag in tagged_tokens if word
17                                         not in stop_words]
18
19     return ' '.join(processed_tokens)
20
21 # Terapkan preprocessing pada kolom "Description" di vacancy_df
vacancy_df_1 = vacancy_df.copy()
vacancy_df_1["Description"] =
vacancy_df["Description"].apply(preprocess_vacancy)
```

Penjelasan dari Kode Program 5.5 mengenai implementasi *preprocessing* isian teks deskripsi kualifikasi lowongan kerja, yaitu:

1. Baris 1 merupakan proses untuk mendefinisikan fungsi `preprocess_vacancy` yang menerima parameter `text` untuk memproses teks deskripsi lowongan kerja.
2. Baris 2 merupakan proses untuk mengubah teks input menjadi huruf kecil menggunakan `text.lower()` untuk standarisasi.
3. Baris 3 merupakan proses untuk menghapus angka-angka dari teks.
4. Baris 4 merupakan proses untuk menghapus semua tanda baca dari teks berdasarkan daftar *punctuation* dari `string.punctuation`.
5. Baris 5 merupakan proses untuk menghapus spasi berlebih dengan memecah teks menjadi kata-kata menggunakan `split()`, lalu menggabungkannya kembali menggunakan `' '.join()` dengan spasi tunggal
6. Baris 7-8 merupakan proses untuk memecah teks menjadi daftar kata (*token*) menggunakan `word_tokenize()` dari NLTK dan menyimpannya ke variabel `tokens`.
7. Baris 10-11 merupakan proses untuk memberikan *part-of-speech* (POS) *tag* pada setiap token menggunakan `pos_tag()` dari NLTK dan menyimpannya ke variabel `tagged_tokens`.
8. Baris 13-15 merupakan proses untuk membuat *list* `processed_tokens` dengan mengiterasi `tagged_tokens`, menghapus kata-kata yang ada di `stop_words`, dan melakukan lematisasi pada setiap kata menggunakan `lemmatizer.lemmatize()` dengan tipe kata dari `get_wordnet_pos(tag)`.

9. Baris 17 merupakan proses untuk menggabungkan token yang telah diproses menjadi satu *string* dengan spasi sebagai pemisah menggunakan '*' .join()*' dan mengembalikannya sebagai hasil dari fungsi *preprocess\_vacancy*.
10. Baris 19-20 merupakan proses untuk membuat salinan DataFrame *vacancy\_df* menggunakan *copy()* dan menyimpannya ke variabel *vacancy\_df\_1* guna menghindari modifikasi data asli.
11. Baris 21 merupakan proses untuk menerapkan fungsi *preprocess\_vacancy* ke kolom *Description* pada DataFrame *vacancy\_df* dan menyimpan hasilnya ke kolom *Description* pada DataFrame *vacancy\_df\_1*.

#### 5.4 Implementasi Kode Program Representasi Teks TF-IDF

Dalam kode program ini, dilakukan implementasi representasi teks menggunakan TF-IDF dengan *TfidfVectorizer* dari *library Scikit-learn* untuk mengubah teks pada kolom *Text* dari *resume\_df\_1* dan kolom *Description* dari *vacancy\_df\_1* menjadi vektor TF-IDF. Vektor-vektor ini disimpan sebagai *list* dalam kolom baru *TFIDF\_Vector* pada kedua DataFrame. Implementasi kode program tertera pada Kode Program 5.6.

##### Kode Program 5.6 Implementasi kode program representasi teks TF-IDF

```

1 # Inisialisasi TF-IDF Vectorizer
2 vectorizer = TfidfVectorizer()
3
4 # Fit dan transform data
5 tfidf_resume = vectorizer.fit_transform(resume_df_1["Text"])
6 tfidf_vacancy =
7 vectorizer.transform(vacancy_df_1["Description"])
8
9 # Simpan vektor dalam kolom baru
10 resume_df_1["TFIDF_Vector"] = list(tfidf_resume.toarray())
    vacancy_df_1["TFIDF_Vector"] = list(tfidf_vacancy.toarray())

```

Penjelasan dari Kode Program 5.6 mengenai implementasi representasi teks dengan pendekatan TF-IDF, yaitu:

1. Baris 1-2 merupakan proses untuk menginisialisasi objek *TfidfVectorizer* dari *library Scikit-learn* menggunakan *TfidfVectorizer()* dan menyimpannya ke variabel *vectorizer* untuk mengubah teks menjadi vektor TF-IDF.
2. Baris 4-5 merupakan proses untuk mempelajari kosa kata dari kolom *Text* pada DataFrame *resume\_df\_1* dan mengubahnya menjadi matriks TF-IDF menggunakan *fit\_transform()*, lalu menyimpan hasilnya ke variabel *tfidf\_resume*.

3. Baris 6 merupakan proses untuk mengubah kolom `Description` pada DataFrame `vacancy_df_1` menjadi matriks TF-IDF menggunakan `transform()` berdasarkan kosa kata yang telah dipelajari dari `resume_df_1["Text"]`, lalu menyimpan hasilnya ke variabel `tfidf_vacancy`.
4. Baris 8-9 merupakan proses untuk mengonversi matriks TF-IDF `tfidf_resume` menjadi `array` menggunakan `toarray()` dan menyimpan setiap vektor sebagai `list` dalam kolom baru `TFIDF_Vector` pada DataFrame `resume_df_1`.
5. Baris 10 merupakan proses untuk mengonversi matriks TF-IDF `tfidf_vacancy` menjadi `array` menggunakan `toarray()` dan menyimpan setiap vektor sebagai `list` dalam kolom baru `TFIDF_Vector` pada DataFrame `vacancy_df_1`.

## 5.5 Implementasi Kode Program Representasi Teks Word2Vec

Dalam kode program ini, dilakukan implementasi representasi teks menggunakan Word2Vec dari *library* Gensim. Kode ini memproses kolom `Text` pada `resume_df_1` dan kolom `Description` pada `vacancy_df_1` untuk pelatihan model Word2Vec dengan menggabungkan teks resume dan deskripsi lowongan kerja. Model Word2Vec dilatih dengan parameter ukuran vektor kata bernilai 100, jarak antar kata dalam konteks (*window*) bernilai 5, menggunakan *skip-gram*, dan *learning rate* bernilai 0,1. Hasil vektor disimpan ke kolom baru `W2V_Vector` pada kedua DataFrame. Implementasi kode program tertera pada Kode Program 5.7.

### Kode Program 5.7 Implementasi kode program representasi teks Word2Vec

```

1  def tokenize_text(text):
2      return word_tokenize(text.lower()) # Tokenisasi dan ubah
3      ke huruf kecil
4
5      # Tokenisasi teks dari resume_df_1 dan vacancy_df_1
6      resume_texts =
7      resume_df_1['Text'].dropna().apply(tokenize_text).tolist()
8      vacancy_texts =
9      vacancy_df_1['Description'].dropna().apply(tokenize_text).tolis
10     t()
11
12     # Gabungkan semua teks untuk pelatihan Word2Vec
13     all_texts = resume_texts + vacancy_texts
14
15     # Latih model Word2Vec
16     word2vec_model = Word2Vec(
17         sentences=all_texts,
18         vector_size=100 #ukuran vektor kata
19         window=5, #jarak maksimum antar kata dalam konteks
20         workers=4, #jumlah thread untuk pelatihan
21         sg=1,
22         alpha=0,1
23     )
24
25     # Menghitung vektor rata-rata dokumen
26     def get_document_vector(text, model):

```

### Kode Program 5.7 Implementasi kode program representasi teks Word2Vec (lanjutan)

```
23     words = tokenize_text(text)
24     word_vectors = [model.wv[word] for word in words if word in
25         model.wv]
26     if not word_vectors: # Jika tidak ada kata yang dikenali
27         return np.zeros(model.vector_size)
28     return np.mean(word_vectors, axis=0)
29
30     # Penerapan ke dataset
31     resume_df_1['W2V_Vector'] = resume_df_1['Text'].apply(lambda x:
32         get_document_vector(x, word2vec_model))
33     vacancy_df_1['W2V_Vector'] =
34     vacancy_df_1['Description'].apply(lambda x:
35         get_document_vector(x, word2vec_model))
```

Penjelasan dari Kode Program 5.7 mengenai implementasi representasi teks dengan pendekatan Word2Vec, yaitu:

1. Baris 1-2 merupakan proses untuk mendefinisikan fungsi `tokenize_text` yang mengambil `text` sebagai input, mengonversi ke huruf kecil (*lower casing*) dengan `text.lower()`, dan melakukan tokenisasi menggunakan `word_tokenize` dari *library* NLTK.
2. Baris 3-4 merupakan proses untuk menghapus nilai kosong dari kolom `Text` pada DataFrame `resume_df_1` menggunakan `dropna()`, melakukan tokenisasi kata dengan `word_tokenize()`, dan mengonversi hasilnya menjadi *list* menggunakan `tolist()`, lalu menyimpannya ke variabel `resume_texts`.
3. Baris 5-6 merupakan proses untuk menghapus nilai kosong dari kolom `Description` pada DataFrame `vacancy_df_1` menggunakan `dropna()`, melakukan tokenisasi kata dengan `word_tokenize()`, dan mengonversi hasilnya menjadi *list* menggunakan `tolist()`, lalu menyimpannya ke variabel `vacancy_texts`.
4. Baris 8-9 merupakan proses untuk menggabungkan *list* `resume_texts` dan `vacancy_texts` menjadi satu *list* `all_texts` untuk digunakan dalam pelatihan model Word2Vec.
5. Baris 11-19 merupakan proses untuk melatih model Word2Vec menggunakan `Word2Vec()` dari *library* Gensim dengan parameter: `sentences=all_texts` (data teks), `vector_size=100` (ukuran vektor kata), `window=5` (jarak konteks kata), `workers=4` (jumlah *thread*), `sg=1` (menggunakan algoritma *skip-gram*), dan `alpha=0,1` (*learning rate*), lalu menyimpan model ke variabel `word2vec_model`.

6. Baris 21-27 merupakan proses untuk mendefinisikan fungsi `get_document_vector` yang menghitung vektor rata-rata dokumen dengan melakukan tokenisasi teks menggunakan `tokenize_text()`, mengambil vektor kata dari `model.wv` untuk kata yang ada di `model`. Fungsi ini mengembalikan vektor nol dengan panjang `vector_size` jika tidak ada kata yang dikenali atau mengembalikan hasil rata-rata vektor kata menggunakan `np.mean()`.
7. Baris 30 merupakan proses untuk menerapkan fungsi `get_document_vector` ke kolom `Text` pada `DataFrame` `resume_df_1` menggunakan `apply()` dengan model `word2vec_model`, lalu menyimpan vektor rata-rata dokumen ke kolom baru `W2V_Vector`.
8. Baris 231 merupakan proses untuk menerapkan fungsi `get_document_vector` ke kolom `Description` pada `DataFrame` `vacancy_df_1` menggunakan `apply()` dengan model `word2vec_model`, lalu menyimpan vektor rata-rata dokumen ke kolom baru `W2V_Vector`.

## 5.6 Implementasi Kode Program Perhitungan Similaritas

### 5.6.1 Implementasi Kode Program *Improved Sqrt-Cosine Similarity*

Dalam kode program ini, dilakukan penghitungan kemiripan antara dua vektor menggunakan pendekatan *Improved Sqrt-Cosine (ISC) Similarity*. Prosesnya mencakup memeriksa validitas input, mengubah elemen vektor menjadi positif, menghitung pembilang (jumlah akar kuadrat perkalian elemen) dan penyebut (perkalian akar kuadrat jumlah elemen), lalu mengembalikan nilai kemiripan atau 0 jika perhitungan tidak valid. Implementasi kode program tertera pada Kode Program 5.8.

#### Kode Program 5.8 Implementasi kode program *Improved Sqrt-Cosine Similarity*

```

1   # Fungsi Improved Sqrt-Cosine Similarity (ISC)
2   def improved_sqrt_cosine_similarity(x, y):
3       if x is None or y is None or len(x) != len(y):
4           return 0
5
6       # Mengambil nilai absolut untuk penggunaan dengan Word2Vec
7       x = np.abs(x)
8       y = np.abs(y)
9
10      # Menghitung pembilang dan penyebut sesuai dengan rumus ISC
11      numerator = np.sum(np.sqrt(x * y))
12      denominator = np.sqrt(np.sum(x)) * np.sqrt(np.sum(y))
13
14      # Menghitung ISC
15      isc = numerator / denominator if denominator != 0 else 0
16      return isc

```

Penjelasan dari Kode Program 5.8 mengenai implementasi perhitungan *Improved Sqrt-Cosine Similarity*, yaitu:

1. Baris 1-2 merupakan proses untuk mendefinisikan fungsi `improved_sqrt_cosine_similarity` yang menerima dua parameter, yakni `x` dan `y` (vektor) untuk menghitung kemiripan menggunakan metrik *Improved Sqrt-Cosine Similarity* (ISC).
2. Baris 3-4 merupakan proses untuk memeriksa apakah salah satu vektor `x` atau `y` adalah `None` atau memiliki panjang yang berbeda menggunakan kondisi `if`. Kemudian, mengembalikan nilai `0` jika kondisi tersebut `true`.
3. Baris 6-8 merupakan proses untuk mengubah semua elemen vektor `x` atau `y` menjadi nilai absolut menggunakan `np.abs()`. Absolut ini digunakan ketika mengimplementasi pendekatan Word2Vec dengan ISC.
4. Baris 10-11 merupakan proses untuk menghitung numerator (pembilang) rumus ISC dengan mengalikan elemen-elemen vektor `x` dan `y`, mengambil akar kuadrat dari hasil perkalian dengan `np.sqrt()`, dan menjumlahkan semua hasilnya menggunakan `np.sum()`.
5. Baris 12 merupakan proses untuk menghitung denominator (penyebut) rumus ISC dengan menjumlahkan elemen vektor `x` atau `y` menggunakan `np.sum()`, mengambil akar kuadrat dari masing-masing jumlah dengan `np.sqrt()`, lalu mengalikan kedua akar tersebut.
6. Baris 14-16 merupakan proses untuk menghitung nilai ISC yang disimpan pada variabel `isc` dengan membagi numerator dengan denominator jika denominator tidak nol atau mengembalikan `0` jika denominator nol untuk menghindari pembagian dengan nol. Kemudian, mengembalikan nilai variabel `isc` yang telah dihitung sebagai hasil dari fungsi `improved_sqrt_cosine_similarity`.

### 5.6.2 Implementasi Kode Program TF-IDF dan *Improved Sqrt-Cosine Similarity*

Dalam kode program ini, dilakukan perhitungan kemiripan antara resume dan kualifikasi lowongan kerja menggunakan *Improved Sqrt-Cosine* (ISC) *Similarity* untuk vektor TF-IDF, dengan dua skenario, yakni “Tanpa Bobot” (hasil kemiripan semua `section` dirata-ratakan) dan “Dengan Bobot” (hasil kemiripan semua `section` diberikan bobot persentase berdasarkan `section_df`). Hasil disimpan dalam DataFrame `result_df_tfidf`, dan lima resume teratas untuk setiap posisi ditampilkan berdasarkan kemiripan “Tanpa Bobot” dan “Dengan Bobot”. Implementasi kode program tertera pada Kode Program 5.9.

#### Kode Program 5.9 Implementasi kode program TF-IDF dan *Improved Sqrt-Cosine Similarity*

```
1 start_time = time.time() # Catat waktu mulai
2
3 # List untuk menyimpan hasil similarity setiap resume
4 final_results_tfidf = []
5
```

## Kode Program 5.9 Implementasi kode program TF-IDF dan *Improved Sqrt-Cosine Similarity* (lanjutan)

```

6      # Looping untuk setiap job vacancy
7      for _, vacancy_row_tfidf in tqdm(vacancy_df_1.iterrows(), 
8          total=len(vacancy_df_1), desc="Processing Vacancies"):
9          vacancy_category_tfidf = vacancy_row_tfidf["Category"]
10         job_vec_tfidf = vacancy_row_tfidf["TFIDF_Vector"]
11         position_name = vacancy_row_tfidf["Position"]
12
13         # Ambil bobot section sesuai kategori dan ubah ke skala
14         # desimal
15         category_weights = (section_df[section_df["Category"] == 
16             vacancy_category_tfidf]
17                         .set_index("Section") ["Bobot"]
18                         .div(100)
19                         .to_dict())
20
21         # Looping untuk setiap resume
22         for resume_id in resume_df_1["ID"].unique():
23             resume_sections = resume_df_1[resume_df_1["ID"] == 
24                 resume_id]
25             similarity_scores_tfidf = {}
26
27             # Hitung similarity untuk setiap section
28             for _, section_row_tfidf in resume_sections.iterrows():
29                 section_name_tfidf = section_row_tfidf["Section"]
30                 section_vec_tfidf =
31                     section_row_tfidf["TFIDF_Vector"]
32
33                 sim_tfidf =
34                     improved_sqrt_cosine_similarity(section_vec_tfidf,
35                         job_vec_tfidf)
36                     similarity_scores_tfidf[section_name_tfidf] =
37                     sim_tfidf
38
39             # Versi 1: Tanpa bobot (rata-rata similarity semua
40             # section)
41             sim_no_weight = sum(similarity_scores_tfidf.values()) / 
42             len(similarity_scores_tfidf) if similarity_scores_tfidf else 0
43
44             # Versi 2: Dengan bobot (weighted sum tanpa normalisasi
45             # total weight)
46             weighted_sum_v2 = 0
47             total_weight_v2 = 0
48             for sec in similarity_scores_tfidf:
49                 sim = similarity_scores_tfidf[sec]
50                 weight = category_weights.get(sec, 0)
51                 weighted_sum_v2 += sim * weight
52                 total_weight_v2 += weight
53             sim_with_weight = weighted_sum_v2 / total_weight_v2 if
54             total_weight_v2 > 0 else 0
55
56             # Simpan hasil
57             final_results_tfidf.append((resume_id, position_name,
58                 sim_no_weight, sim_with_weight))
59
60             # Catat waktu selesai dan hitung durasi
61             end_time = time.time()
62             total_time = end_time - start_time
63
64             # Konversi waktu ke format yang lebih mudah dibaca
65             minutes = int(total_time // 60)
66             seconds = int(total_time % 60)

```

### Kode Program 5.9 Implementasi kode program TF-IDF dan *Improved Sqrt-Cosine Similarity* (lanjutan)

```
54
55     # Buat DataFrame hasil similarity
56     result_df_tfidf = pd.DataFrame(
57         final_results_tfidf,
58         columns=["Resume_ID", "Position", "Similarity_No_Weight",
59         "Similarity_With_Weight"]
60     )
61
62     # Tampilkan waktu total
63     print(f"Total waktu pemrosesan: {minutes} menit {seconds}
64     detik")
65
66     print("Similaritas Tanpa Bobot Section")
67
68     top5_per_position_no_weight =
69     result_df_tfidf.groupby('Position', group_keys=False).apply(
70         lambda x: x.nlargest(5, 'Similarity_No_Weight')
71     )
72
73     grouped_no_weight =
74     top5_per_position_no_weight.groupby('Position')
75     for position, group in grouped_no_weight:
76         print(f"\nPosition: {position}")
77         print(group[['Resume_ID', 'Similarity_No_Weight']])
78
79     print("Similaritas Dengan Bobot Section")
80
81     top5_per_position_with_weight =
82     result_df_tfidf.groupby('Position', group_keys=False).apply(
83         lambda x: x.nlargest(5, 'Similarity_With_Weight')
84     )
85
86     grouped_with_weight =
87     top5_per_position_with_weight.groupby('Position')
88     for position, group in grouped_with_weight:
89         print(f"\nPosition: {position}")
90         print(group[['Resume_ID', 'Similarity_With_Weight']])
```

Penjelasan dari Kode Program 5.9 mengenai implementasi perhitungan similaritas antara resume dan kualifikasi lowongan kerja dengan *Improved Sqrt-Cosine* (ISC) *Similarity* jika menggunakan vektor TF-IDF, yaitu:

1. Baris 1 merupakan proses untuk mencatat waktu mulai eksekusi menggunakan `time.time()` dan menyimpannya ke variabel `start_time`.
2. Baris 3-4 merupakan proses untuk membuat *list* kosong `final_results_tfidf` untuk menyimpan hasil perhitungan kemiripan (*similarity*) antara resume dan kualifikasi lowongan kerja.
3. Baris 6-7 merupakan proses untuk memulai iterasi melalui setiap baris di DataFrame `vacancy_df_1` menggunakan `iterrows()` dengan *progress bar* dari `tqdm` untuk menampilkan kemajuan pemrosesan.
4. Baris 8 merupakan pengambilan nilai kolom `Category` dari baris lowongan kerja pada iterasi terkini dan menyimpannya ke variabel `vacancy_category_tfidf`.

5. Baris 9 merupakan pengambilan vektor TF-IDF dari kolom `TFIDF_Vector` pada baris lowongan kerja di iterasi terkini dan menyimpannya ke variabel `job_vec_tfidf`.
6. Baris 10 merupakan pengambilan nilai kolom `Position` dari baris lowongan kerja pada iterasi terkini dan menyimpannya ke variabel `position_name`.
7. Baris 12-16 merupakan proses untuk memfilter DataFrame `section_df` berdasarkan `vacancy_category_tfidf`, menetapkan kolom `Section` sebagai indeks, mengambil kolom `Bobot`, membaginya dengan 100 menggunakan `div(100)` untuk mengubah ke skala desimal (karena bobot dalam persentase), dan mengonversinya menjadi `dictionary` menggunakan `to_dict()`, lalu menyimpannya ke variabel `category_weights` sebagai persentase bobot per `section` untuk kualifikasi lowongan kerja pada iterasi terkini sesuai industrinya.
8. Baris 18-19 merupakan proses untuk memulai iterasi melalui setiap nilai unik di kolom `ID` pada DataFrame `resume_df_1` menggunakan `unique()`.
9. Baris 20 merupakan proses untuk memfilter DataFrame `resume_df_1` untuk mendapatkan semua baris dengan kolom `ID` yang sesuai dengan iterasi `resume_id` terkini dan menyimpannya ke variabel `resume_sections`.
10. Baris 21 merupakan pembuatan `dictionary` kosong `similarity_scores_tfidf` untuk menyimpan skor kemiripan setiap section dalam `resume`.
11. Baris 23-24 merupakan proses untuk memulai iterasi melalui setiap baris di `resume_sections` menggunakan `iterrows()` untuk memproses setiap `section` dalam `resume`.
12. Baris 25 merupakan proses untuk mengambil nilai kolom `Section` dari baris `section` terkini dan menyimpannya ke variabel `section_name_tfidf`.
13. Baris 26 merupakan proses untuk mengambil vektor TF-IDF dari kolom `TFIDF_Vector` pada baris `section` saat ini dan menyimpannya ke variabel `section_vec_tfidf`.
14. Baris 28 merupakan proses untuk menghitung kemiripan antara vektor `section` pada `resume` (`section_vec_tfidf`) dan vektor kualifikasi lowongan kerja (`job_vec_tfidf`) menggunakan fungsi `improved_sqrt_cosine_similarity`, lalu menyimpan hasilnya ke variabel `sim_tfidf`.
15. Baris 29 merupakan proses untuk menyimpan skor kemiripan `sim_tfidf` ke `dictionary similarity_scores_tfidf`.

16. Baris 31-32 merupakan proses untuk menguji skenario "Tanpa Bobot" dengan menghitung rata-rata kemiripan dengan menjumlahkan semua skor kemiripan di `similarity_scores_tfidf` menggunakan `sum()` dan membaginya dengan jumlah `section` menggunakan `len()`, atau mengembalikan 0 jika `dictionary` kosong, lalu menyimpan hasilnya ke variabel `sim_no_weight`.
17. Baris 34-42 merupakan proses untuk menguji skenario "Dengan Bobot", dimana baris 35-36 menginisialisasi variabel `weighted_sum_v2` dan `total_weight_v2` dengan nilai 0 untuk menghitung jumlah kemiripan terbobot dan total bobot.
18. Baris 37-41 merupakan proses untuk mengiterasi setiap `section` di `similarity_scores_tfidf`, mengambil skor kemiripan yang disimpan pada variabel `sim`, mendapatkan bobot `section` dari variabel `category_weights` dengan `default 0` menggunakan `get()`, mengalikan skor "Dengan Bobot" untuk menambah ke `weighted_sum_v2`, dan menambah bobot ke `total_weight_v2`.
19. Baris 42 merupakan proses untuk menghitung total `similarity` pada skenario "Dengan Bobot" dengan membagi `weighted_sum_v2` dengan `total_weight_v2` jika `total_weight_v2` lebih dari 0 atau mengembalikan nilai 0 jika tidak. Kemudian, menyimpan hasilnya ke variabel `sim_with_weight`.
20. Baris 44-45 merupakan proses untuk menambahkan data `resume_id`, `position_name`, `sim_no_weight`, dan `sim_with_weight` ke `list final_results_tfidf` menggunakan `append()`.
21. Baris 47-49 merupakan proses untuk mencatat waktu selesai menggunakan `time.time()`, menghitung durasi dengan mengurangkan `start_time` dari `end_time`, dan menyimpan hasilnya ke variabel `total_time`.
22. Baris 51-53 merupakan proses untuk mengonversi `total_time` ke menit dan detik, lalu menyimpannya ke variabel `minutes` dan `seconds`.
23. Baris 55-59 merupakan proses untuk membuat DataFrame `result_df_tfidf` dari `list final_results_tfidf` menggunakan `pd.DataFrame()`, dengan kolom `Resume_ID`, `Position`, `Similarity_No_Weight`, dan `Similarity_With_Weight`.
24. Baris 61-62 merupakan proses untuk mencetak total waktu pemrosesan dalam format menit dan detik menggunakan `print()`.
25. Baris 64 merupakan proses untuk mencetak judul "Similaritas Tanpa Bobot Section" menggunakan `print()`.
26. Baris 66-68 merupakan proses untuk mengelompokkan DataFrame `result_df_tfidf` berdasarkan kolom `Position` menggunakan `groupby()`, lalu memilih 5 baris dengan nilai `Similarity_No_Weight` tertinggi untuk setiap posisi menggunakan `nlargest()`, dan menyimpan hasilnya ke variabel `top5_per_position_no_weight`.

27. Baris 70-73 merupakan proses untuk mengelompokkan top5\_per\_position\_no\_weight berdasarkan kolom Position, mengiterasi setiap kualifikasi lowongan kerja (Position), mencetak nama posisi, mencetak nilai kolom Resume\_ID, serta mencetak nilai Similarity\_No\_Weight untuk setiap kualifikasi lowongan kerja.
  28. Baris 75 merupakan proses untuk mencetak judul "Similaritas Dengan Bobot Section" menggunakan print().
  29. Baris 77-79 merupakan proses untuk mengelompokkan DataFrame result\_df\_tfidf berdasarkan kolom Position menggunakan groupby(), lalu memilih 5 baris dengan nilai Similarity\_With\_Weight tertinggi untuk setiap posisi menggunakan nlargest(), dan menyimpan hasilnya ke variabel top5\_per\_position\_with\_weight.
  30. Baris 81-84 merupakan proses untuk mengelompokkan top5\_per\_position\_with\_weight berdasarkan Position, mengiterasi setiap kualifikasi lowongan kerja (Position), mencetak nama posisi, mencetak kolom Resume\_ID, serta Similarity\_With\_Weight untuk setiap kualifikasi lowongan kerja.

### 5.6.3 Implementasi Kode Program Word2Vec dan *Cosine Similarity*

Dalam kode program ini, dilakukan menghitung kemiripan antara resume dan kualifikasi lowongan kerja menggunakan metrik *Cosine Similarity* (CosSim) untuk vektor Word2Vec, dengan dua skenario, yakni “Tanpa Bobot” (hasil kemiripan semua *section* dirata-ratakan) dan “Dengan Bobot” (hasil kemiripan semua *section* diberikan bobot persentase berdasarkan *section\_df*). Hasil disimpan dalam DataFrame `result_df_w2v`, dan lima resume teratas untuk setiap posisi ditampilkan berdasarkan kemiripan “Tanpa Bobot” dan “Dengan Bobot”. Implementasi kode program tertera pada Kode Program 5.10.

## Kode Program 5.10 Implementasi kode program Word2Vec dan *Cosine Similarity*

```
1 start_time = time.time() # Catat waktu mulai
2
3 # List untuk menyimpan hasil
4 final_results_w2v = []
5
6 # Iterasi untuk setiap vacancy
7 for vac_idx, vac_row in tqdm(vacancy_df_1.iterrows(),
8 total=len(vacancy_df_1), desc="Processing Vacancies"):
9     position = vac_row['Position']
10    vacancy_category = vac_row['Category']
11    job_vec_w2v = np.array([vac_row['W2V_Vector']]) # Vektor
12    vacancy dalam bentuk 2D untuk cosine_similarity
13
14    # Ambil bobot section sesuai kategori dan ubah ke skala
15    # desimal
16    category_weights = (section_df[section_df["Category"] ==
17                          vacancy_category]
18                           .set_index("Section")["Bobot"]
19                           .div(100))
```

## Kode Program 5.10 Implementasi kode program Word2Vec dan Cosine Similarity (lanjutan)

```
16             .to_dict())
17
18     # Iterasi untuk setiap resume
19     for resume_id in resume_df_1['ID'].unique():
20         resume_sections = resume_df_1[resume_df_1['ID'] == resume_id]
21         similarity_scores = {}
22
23         # Hitung similarity untuk setiap section
24         for _, section_row in resume_sections.iterrows():
25             section_name = section_row['Section']
26             section_vector =
27             np.array([section_row['W2V_Vector']]) # Vektor section dalam bentuk 2D
28
29             # Hitung cosine similarity antara section dan vacancy
30             sim_score = cosine_similarity(section_vector,
31             job_vec_w2v)[0][0]
32             similarity_scores[section_name] = sim_score
33
34             # Versi 1: Tanpa bobot (rata-rata similarity semua section)
35             sim_no_weight = sum(similarity_scores.values()) / len(similarity_scores) if similarity_scores else 0
36
37             # Versi 2: Dengan bobot (weighted sum tanpa normalisasi ketat)
38             weighted_sum_v2 = 0
39             total_weight_v2 = 0
40             for section in similarity_scores:
41                 sim = similarity_scores.get(section, 0)
42                 weight = category_weights.get(section, 0)
43                 weighted_sum_v2 += sim * weight
44                 total_weight_v2 += weight
45             sim_with_weight = weighted_sum_v2 / total_weight_v2 if total_weight_v2 > 0 else 0
46
47             # Simpan hasil
48             final_results_w2v.append({
49                 'Resume_ID': resume_id,
50                 'Position': position,
51                 'Similarity_No_Weight': sim_no_weight,
52                 'Similarity_With_Weight': sim_with_weight,
53             })
54
55     # Catat waktu selesai dan hitung durasi
56     end_time = time.time()
57     total_time = end_time - start_time
58
59     # Konversi waktu ke format yang lebih mudah dibaca
60     minutes = int(total_time // 60)
61     seconds = int(total_time % 60)
62
63     # Konversi hasil ke DataFrame
64     result_df_w2v = pd.DataFrame(final_results_w2v)
65
66     # Tampilkan waktu total
67     print(f"Total waktu pemrosesan: {minutes} menit {seconds} detik")
```

### Kode Program 5.10 Implementasi kode program Word2Vec dan Cosine Similarity (lanjutan)

```
67     print("Similaritas Tanpa Bobot Section")
68
69     top5_per_position_no_weight = result_df_w2v.groupby('Position',
70             group_keys=False).apply(
71         lambda x: x.nlargest(5, 'Similarity_No_Weight')
72     )
73
73     grouped_no_weight =
74     top5_per_position_no_weight.groupby('Position')
75     for position, group in grouped_no_weight:
76         print(f"\nPosition: {position}")
76         print(group[['Resume_ID', 'Similarity_No_Weight']])
77
78     print("Similaritas Dengan Bobot Section")
79
80     top5_per_position_with_weight =
81     result_df_w2v.groupby('Position', group_keys=False).apply(
82         lambda x: x.nlargest(5, 'Similarity_With_Weight')
83     )
84
84     grouped_with_weight =
85     top5_per_position_with_weight.groupby('Position')
86     for position, group in grouped_with_weight:
86         print(f"\nPosition: {position}")
86         print(group[['Resume_ID', 'Similarity_With_Weight']])
87
```

Penjelasan dari Kode Program 5.10 mengenai implementasi perhitungan similaritas antara resume dan kualifikasi lowongan kerja dengan *Cosine Similarity* (CosSim) jika menggunakan vektor Word2Vec, yaitu:

1. Baris 1 merupakan proses untuk mencatat waktu mulai eksekusi menggunakan `time.time()` dan menyimpannya ke variabel `start_time`.
2. Baris 3-4 merupakan proses untuk membuat *list* kosong `final_results_w2v` untuk menyimpan hasil perhitungan kemiripan (*similarity*) antara resume dan kualifikasi lowongan kerja.
3. Baris 6-7 merupakan proses untuk memulai iterasi melalui setiap baris di DataFrame `vacancy_df_1` menggunakan `iterrows()` dengan *progress bar* dari `tqdm` untuk menampilkan kemajuan pemrosesan.
4. Baris 8 merupakan proses untuk mengambil nilai kolom `Position` dari baris kualifikasi lowongan kerja terkini dan menyimpannya ke variabel `position`.
5. Baris 9 merupakan pengambilan nilai kolom `Category` dari baris lowongan kerja pada iterasi terkini dan menyimpannya ke variabel `vacancy_category`.
6. Baris 10 merupakan pengambilan vektor Word2Vec dari kolom `W2V_Vector` pada baris kualifikasi lowongan kerja di iterasi terkini dan menyimpannya ke variabel `job_vec_w2v`.

7. Baris 12-16 merupakan proses untuk memfilter DataFrame `section_df` berdasarkan `vacancy_category`, menetapkan kolom `Section` sebagai indeks, mengambil kolom `Bobot`, membaginya dengan 100 menggunakan `div(100)` untuk mengubah ke skala desimal (karena bobot dalam persentase), dan mengonversinya menjadi `dictionary` menggunakan `to_dict()`, lalu menyimpannya ke variabel `category_weights` sebagai persentase bobot per `section` untuk kualifikasi lowongan kerja pada iterasi terkini sesuai industrinya.
8. Baris 18-19 merupakan proses untuk memulai iterasi melalui setiap nilai unik di kolom `ID` pada DataFrame `resume_df_1` menggunakan `unique()`.
9. Baris 20 merupakan proses untuk memfilter DataFrame `resume_df_1` untuk mendapatkan semua baris dengan kolom `ID` yang sesuai dengan iterasi `resume_id` terkini dan menyimpannya ke variabel `resume_sections`.
10. Baris 21 merupakan pembuatan `dictionary` kosong `similarity_scores` untuk menyimpan skor kemiripan setiap `section` dalam `resume`.
11. Baris 23-24 merupakan proses untuk memulai iterasi melalui setiap baris di `resume_sections` menggunakan `iterrows()` untuk memproses setiap `section` dalam `resume`.
12. Baris 25 merupakan proses untuk mengambil nilai kolom `Section` dari baris `section` terkini dan menyimpannya ke variabel `section_name`.
13. Baris 26 merupakan proses untuk mengambil vektor Word2Vec dari kolom `W2V_Vector` pada baris `section` saat ini dan menyimpannya ke variabel `section_vector`.
14. Baris 28-29 merupakan proses untuk menghitung kemiripan antara vektor `section` pada `resume` (`section_vector`) dan vektor kualifikasi lowongan kerja (`vac_vector`) menggunakan fungsi `cosine_similarity` dari library Scikit-learn, lalu menyimpan hasilnya ke variabel `sim_score`.
15. Baris 30 merupakan proses untuk menyimpan skor kemiripan `sim_score` ke `dictionary similarity_scores`.
16. Baris 32-33 merupakan proses untuk menguji skenario “Tanpa Bobot” dengan menghitung rata-rata kemiripan dengan menjumlahkan semua skor kemiripan di `similarity_scores` menggunakan `sum()` dan membaginya dengan jumlah `section` menggunakan `len()`, atau mengembalikan 0 jika `dictionary` kosong, lalu menyimpan hasilnya ke variabel `sim_no_weight`.
17. Baris 35-43 merupakan proses untuk menguji skenario “Dengan Bobot”, dimana baris 36-37 menginisialisasi variabel `weighted_sum_v2` dan `total_weight_v2` dengan nilai 0 untuk menghitung jumlah kemiripan terbobot dan total bobot.

18. Baris 38-42 merupakan proses untuk mengiterasi setiap *section* di *similarity\_scores*, mengambil skor kemiripan yang disimpan pada variabel *sim*, mendapatkan bobot *section* dari variabel *category\_weights* dengan *default* 0 menggunakan *get()*, mengalikan skor "Dengan Bobot" untuk menambah ke *weighted\_sum\_v2*, dan menambah bobot ke *total\_weight\_v2*.
19. Baris 43 merupakan proses untuk menghitung total *similarity* pada skenario "Dengan Bobot" dengan membagi *weighted\_sum\_v2* dengan *total\_weight\_v2* jika *total\_weight\_v2* lebih dari 0 atau mengembalikan nilai 0 jika tidak. Kemudian, menyimpan hasilnya ke variabel *sim\_with\_weight*.
20. Baris 45-46 merupakan proses untuk menambahkan *data resume\_id*, *position*, *sim\_no\_weight*, dan *sim\_with\_weight* ke *list final\_results\_w2v* menggunakan *append()*.
21. Baris 53-55 merupakan proses untuk mencatat waktu selesai menggunakan *time.time()*, menghitung durasi dengan mengurangkan *start\_time* dari *end\_time*, dan menyimpan hasilnya ke variabel *total\_time*.
22. Baris 57-59 merupakan proses untuk mengonversi *total\_time* ke menit dan detik, lalu menyimpannya ke variabel *minutes* dan *seconds*.
23. Baris 61-62 merupakan proses untuk membuat DataFrame *result\_df\_w2v* dari *list final\_results\_w2v* menggunakan *pd.DataFrame()*, dengan kolom *Resume\_ID*, *Position*, *Similarity\_No\_Weight*, dan *Similarity\_With\_Weight*.
24. Baris 64-65 merupakan proses untuk mencetak total waktu pemrosesan dalam format menit dan detik menggunakan *print()*.
25. Baris 67 merupakan proses untuk mencetak judul "Similaritas Tanpa Bobot Section" menggunakan *print()*.
26. Baris 69-71 merupakan proses untuk mengelompokkan DataFrame *result\_df\_w2v* berdasarkan kolom *Position* menggunakan *groupby()*, lalu memilih 5 baris dengan nilai *Similarity\_No\_Weight* tertinggi untuk setiap posisi menggunakan *nlargest()*, dan menyimpan hasilnya ke variabel *top5\_per\_position\_no\_weight*.
27. Baris 73-76 merupakan proses untuk mengelompokkan *top5\_per\_position\_no\_weight* berdasarkan kolom *Position*, mengiterasi setiap kualifikasi lowongan kerja (*Position*), mencetak nama posisi, mencetak nilai kolom *Resume\_ID*, serta mencetak nilai *Similarity\_No\_Weight* untuk setiap kualifikasi lowongan kerja.
28. Baris 78 merupakan proses untuk mencetak judul "Similaritas Dengan Bobot Section" menggunakan *print()*.

29. Baris 80-82 merupakan proses untuk mengelompokkan DataFrame `result_df_w2v` berdasarkan kolom `Position` menggunakan `groupby()`, lalu memilih 5 baris dengan nilai `Similarity_With_Weight` tertinggi untuk setiap posisi menggunakan `nlargest()`, dan menyimpan hasilnya ke variabel `top5_per_position_with_weight`.
30. Baris 84-87 merupakan proses untuk mengelompokkan `top5_per_position_with_weight` berdasarkan `Position`, mengiterasi setiap kualifikasi lowongan kerja (`Position`), mencetak nama posisi, mencetak kolom `Resume_ID`, serta `Similarity_With_Weight` untuk setiap kualifikasi lowongan kerja.

#### **5.6.4 Implementasi Kode Program Word2Vec dan *Improved Sqrt-Cosine Similarity***

Dalam kode program ini, dilakukan kemiripan antara resume dan kualifikasi lowongan kerja menggunakan metrik *Improved Sqrt-Cosine* (ISC) *Similarity* untuk vektor Word2Vec, dengan dua skenario, yakni “Tanpa Bobot” (hasil kemiripan semua `section` dirata-ratakan) dan “Dengan Bobot” (hasil kemiripan semua `section` diberikan bobot persentase berdasarkan `section_df`). Hasil disimpan dalam DataFrame `result_df_w2v_isc`, dan lima resume teratas untuk setiap posisi ditampilkan berdasarkan kemiripan “Tanpa Bobot” dan ‘Dengan Bobot’. Implementasi kode program tertera pada Kode Program 5.11.

#### **Kode Program 5.11 Implementasi kode program Word2Vec dan *Improved Sqrt-Cosine Similarity***

```

1 start_time = time.time() # Catat waktu mulai
2
3 # List untuk menyimpan hasil
4 final_results_w2v_isc = []
5
6 # Iterasi untuk setiap vacancy
7 for vac_idx, vac_row in tqdm(vacancy_df_1.iterrows(),
8 total=len(vacancy_df_1), desc="Processing Vacancies"):
9     position = vac_row['Position']
10    vacancy_category = vac_row['Category']
11    job_vec_w2v = vac_row['W2V_Vector'] # Vektor vacancy dalam
bentuk 1D
12
13    # Ambil bobot section sesuai kategori dan ubah ke skala
desimal
14    category_weights = (section_df[section_df["Category"] ==
vacancy_category]
15                            .set_index("Section") ["Bobot"]
16                            .div(100)
17                            .to_dict())
18
19    # Iterasi untuk setiap resume
20    for resume_id in resume_df_1['ID'].unique():
21        resume_sections = resume_df_1[resume_df_1['ID'] ==
resume_id]
22        similarity_scores = {}
23
24        # Hitung similarity untuk setiap section

```

## Kode Program 5.11 Implementasi kode program Word2Vec dan *Improved Sqrt-Cosine Similarity* (lanjutan)

```

24         for _, section_row in resume_sections.iterrows():
25             section_name = section_row['Section']
26             section_vector = section_row['W2V_Vector'] # Vektor section dalam bentuk 1D
27
28             # Hitung improved sqrt-cosine similarity antara section dan vacancy
29             sim_score =
30             improved_sqrt_cosine_similarity(section_vector, job_vec_w2v)
31             similarity_scores[section_name] = sim_score
32
33             # Versi 1: Tanpa bobot (rata-rata similarity semua section)
34             sim_no_weight = sum(similarity_scores.values()) /
35             len(similarity_scores) if similarity_scores else 0
36
37             # Versi 2: Dengan bobot (weighted sum tanpa normalisasi ketat)
38             weighted_sum_v2 = 0
39             total_weight_v2 = 0
40             for section in similarity_scores:
41                 sim = similarity_scores.get(section, 0)
42                 weight = category_weights.get(section, 0)
43                 weighted_sum_v2 += sim * weight
44                 total_weight_v2 += weight
45                 sim_with_weight = weighted_sum_v2 / total_weight_v2 if
46                 total_weight_v2 > 0 else 0
47
48             # Simpan hasil
49             final_results_w2v_isc.append({
50                 'Resume_ID': resume_id,
51                 'Position': position,
52                 'Similarity_No_Weight': sim_no_weight,
53                 'Similarity_With_Weight': sim_with_weight,
54             })
55
56
57             # Catat waktu selesai dan hitung durasi
58             end_time = time.time()
59             total_time = end_time - start_time
60
61             # Konversi waktu ke format yang lebih mudah dibaca
62             minutes = int(total_time // 60)
63             seconds = int(total_time % 60)
64
65             # Konversi hasil ke DataFrame
66             result_df_w2v_isc = pd.DataFrame(final_results_w2v_isc)
67
68             # Tampilkan waktu total
69             print(f"Total waktu pemrosesan: {minutes} menit {seconds}
70             detik")
71
72             print("Similaritas Tanpa Bobot Section")
73
74             top5_per_position_no_weight =
75             result_df_w2v_isc.groupby('Position', group_keys=False).apply(
76                 lambda x: x.nlargest(5, 'Similarity_No_Weight')
77             )
78
79             grouped_no_weight =
80             top5_per_position_no_weight.groupby('Position')
81             for position, group in grouped_no_weight:
82
83
84
85
86
87
88
89
90
91
92
93
94

```

### Kode Program 5.11 Implementasi kode program Word2Vec dan *Improved Sqrt-Cosine Similarity* (lanjutan)

```
75     print(f"\nPosition: {position}")
76     print(group[['Resume_ID', 'Similarity_No_Weight']])
77
78     print("Similaritas Dengan Bobot Section")
79
80     top5_per_position_with_weight =
81         result_df_w2v_isc.groupby('Position', group_keys=False).apply(
82             lambda x: x.nlargest(5, 'Similarity_With_Weight')
83         )
84
85     grouped_with_weight =
86         top5_per_position_with_weight.groupby('Position')
87     for position, group in grouped_with_weight:
88         print(f"\nPosition: {position}")
89         print(group[['Resume_ID', 'Similarity_With_Weight']])
```

Penjelasan dari Kode Program 5.11 mengenai implementasi perhitungan similaritas antara resume dan kualifikasi lowongan kerja dengan *Improved Sqrt-Cosine (ISC) Similarity* jika menggunakan vektor Word2Vec, yaitu:

1. Baris 1 merupakan proses untuk mencatat waktu mulai eksekusi menggunakan `time.time()` dan menyimpannya ke variabel `start_time`.
2. Baris 3-4 merupakan proses untuk membuat *list* kosong `final_results_w2v_isc` untuk menyimpan hasil perhitungan kemiripan (*similarity*) antara resume dan kualifikasi lowongan kerja.
3. Baris 6-7 merupakan proses untuk memulai iterasi melalui setiap baris di DataFrame `vacancy_df_1` menggunakan `iterrows()` dengan *progress bar* dari `tqdm` untuk menampilkan kemajuan pemrosesan.
4. Baris 8 merupakan proses untuk mengambil nilai kolom `Position` dari baris kualifikasi lowongan kerja terkini dan menyimpannya ke variabel `position`.
5. Baris 9 merupakan pengambilan nilai kolom `Category` dari baris lowongan kerja pada iterasi terkini dan menyimpannya ke variabel `vacancy_category`.
6. Baris 10 merupakan pengambilan vektor Word2Vec dari kolom `W2V_Vector` pada baris kualifikasi lowongan kerja di iterasi terkini dan menyimpannya ke variabel `job_vec_w2v`.
7. Baris 12-16 merupakan proses untuk memfilter DataFrame `section_df` berdasarkan `vacancy_category`, menetapkan kolom `Section` sebagai indeks, mengambil kolom `Bobot`, membaginya dengan 100 menggunakan `div(100)` untuk mengubah ke skala desimal (karena bobot dalam persentase), dan mengonversinya menjadi *dictionary* menggunakan `to_dict()`, lalu menyimpannya ke variabel `category_weights` sebagai persentase bobot per `section` untuk kualifikasi lowongan kerja pada iterasi terkini sesuai industrinya.

8. Baris 18-19 merupakan proses untuk memulai iterasi melalui setiap nilai unik di kolom ID pada DataFrame `resume_df_1` menggunakan `unique()`.
9. Baris 20 merupakan proses untuk memfilter DataFrame `resume_df_1` untuk mendapatkan semua baris dengan kolom ID yang sesuai dengan iterasi `resume_id` terkini dan menyimpannya ke variabel `resume_sections`.
10. Baris 21 merupakan pembuatan `dictionary` kosong `similarity_scores` untuk menyimpan skor kemiripan setiap section dalam `resume`.
11. Baris 23-24 merupakan proses untuk memulai iterasi melalui setiap baris di `resume_sections` menggunakan `iterrows()` untuk memproses setiap `section` dalam `resume`.
12. Baris 25 merupakan proses untuk mengambil nilai kolom `Section` dari baris `section` terkini dan menyimpannya ke variabel `section_name`.
13. Baris 26 merupakan proses untuk mengambil vektor Word2Vec dari kolom `W2V_Vector` pada baris `section` saat ini dan menyimpannya ke variabel `section_vector`.
14. Baris 28-29 merupakan proses untuk menghitung kemiripan antara vektor `section` pada `resume` (`section_vector`) dan vektor kualifikasi lowongan kerja (`vac_vector`) menggunakan fungsi `cosine_similarity` dari library Scikit-learn, lalu menyimpan hasilnya ke variabel `sim_score`.
15. Baris 30 merupakan proses untuk menyimpan skor kemiripan `sim_score` ke `dictionary similarity_scores`.
16. Baris 32-33 merupakan proses untuk menguji skenario “Tanpa Bobot” dengan menghitung rata-rata kemiripan dengan menjumlahkan semua skor kemiripan di `similarity_scores` menggunakan `sum()` dan membaginya dengan jumlah `section` menggunakan `len()`, atau mengembalikan 0 jika `dictionary` kosong, lalu menyimpan hasilnya ke variabel `sim_no_weight`.
17. Baris 35-43 merupakan proses untuk menguji skenario “Dengan Bobot”, dimana baris 36-37 menginisialisasi variabel `weighted_sum_v2` dan `total_weight_v2` dengan nilai 0 untuk menghitung jumlah kemiripan terbobot dan total bobot.
18. Baris 38-42 merupakan proses untuk mengiterasi setiap `section` di `similarity_scores`, mengambil skor kemiripan yang disimpan pada variabel `sim`, mendapatkan bobot `section` dari variabel `category_weights` dengan `default 0` menggunakan `get()`, mengalikan skor “Dengan Bobot” untuk menambah ke `weighted_sum_v2`, dan menambah bobot ke `total_weight_v2`.

19. Baris 43 merupakan proses untuk menghitung total *similarity* pada skenario "Dengan Bobot" dengan membagi `weighted_sum_v2` dengan `total_weight_v2` jika `total_weight_v2` lebih dari 0 atau mengembalikan nilai 0 jika tidak. Kemudian, menyimpan hasilnya ke variabel `sim_with_weight`.
20. Baris 45-46 merupakan proses untuk menambahkan data `resume_id`, `position`, `sim_no_weight`, dan `sim_with_weight` ke `list final_results_w2v_isc` menggunakan `append()`.
21. Baris 53-55 merupakan proses untuk mencatat waktu selesai menggunakan `time.time()`, menghitung durasi dengan mengurangkan `start_time` dari `end_time`, dan menyimpan hasilnya ke variabel `total_time`.
22. Baris 57-59 merupakan proses untuk mengonversi `total_time` ke menit dan detik, lalu menyimpannya ke variabel `minutes` dan `seconds`.
23. Baris 61-62 merupakan proses untuk membuat DataFrame `result_df_w2v_isc` dari `list final_results_w2v_isc` menggunakan `pd.DataFrame()`, dengan kolom `Resume_ID`, `Position`, `Similarity_No_Weight`, dan `Similarity_With_Weight`.
24. Baris 64-65 merupakan proses untuk mencetak total waktu pemrosesan dalam format menit dan detik menggunakan `print()`.
25. Baris 67 merupakan proses untuk mencetak judul "Similaritas Tanpa Bobot Section" menggunakan `print()`.
26. Baris 69-71 merupakan proses untuk mengelompokkan DataFrame `result_df_w2v_isc` berdasarkan kolom `Position` menggunakan `groupby()`, lalu memilih 5 baris dengan nilai `Similarity_No_Weight` tertinggi untuk setiap posisi menggunakan `nlargest()`, dan menyimpan hasilnya ke variabel `top5_per_position_no_weight`.
27. Baris 73-76 merupakan proses untuk mengelompokkan `top5_per_position_no_weight` berdasarkan kolom `Position`, mengiterasi setiap kualifikasi lowongan kerja (`Position`), mencetak nama posisi, mencetak nilai kolom `Resume_ID`, serta mencetak nilai `Similarity_No_Weight` untuk setiap kualifikasi lowongan kerja.
28. Baris 78 merupakan proses untuk mencetak judul "Similaritas Dengan Bobot Section" menggunakan `print()`.
29. Baris 80-82 merupakan proses untuk mengelompokkan DataFrame `result_df_w2v_isc` berdasarkan kolom `Position` menggunakan `groupby()`, lalu memilih 5 baris dengan nilai `Similarity_With_Weight` tertinggi untuk setiap posisi menggunakan `nlargest()`, dan menyimpan hasilnya ke variabel `top5_per_position_with_weight`.

30. Baris 84-87 merupakan proses untuk mengelompokkan `top5_per_position_with_weight` berdasarkan `Position`, mengiterasi setiap kualifikasi lowongan kerja (`Position`), mencetak nama posisi, mencetak kolom `Resume_ID`, serta `Similarity_With_Weight` untuk setiap kualifikasi lowongan kerja.

## 5.7 Implementasi Kode Program Pengujian

### 5.7.1 Implementasi Kode Program Perhitungan SRCC

Dalam kode program ini, dilakukan pemuatan enam DataFrame dengan format CSV yang berisi peringkat resume dengan dan “Tanpa Bobot” untuk metode TF-IDF dengan *Improved Sqrt-Cosine* (ISC) *Similarity*, Word2Vec dengan *Cosine Similarity* (CosSim), dan Word2Vec dengan ISC. Fungsi `calculate_srcc` menghitung *Spearman Rank Correlation Coefficient* (SRCC) untuk membandingkan peringkat hasil implementasi metode dengan peringkat pakar per posisi. Hasil SRCC digabungkan ke dalam DataFrame `merged_df` dengan kolom `Position` sebagai indeks dan divisualisasikan melalui DataFrame `styled_df` dengan pewarnaan berdasarkan nilai SRCC, ditandai dengan warna hijau jika nilai SRCC pada suatu kualifikasi lowongan kerja (`Position`) bernilai di atas ambang batas dan ditandai dengan warna merah jika nilai SRCC di bawah ambang batas. Implementasi kode program tertera pada Kode Program 5.12.

#### Kode Program 5.12 Implementasi kode program perhitungan *Spearman Rank Correlation Coefficient*

```

1 # Input DataFrame
2 result_df_bobot_tfidf =
3 pd.read_csv(r'C:\Users\mit.itsupport\Downloads\archive2024\Resume2-23\rank_df_bobot_tfidf.csv')
4 result_df_bobot_w2v =
5 pd.read_csv(r'C:\Users\mit.itsupport\Downloads\archive2024\Resume2-23\rank_df_bobot_w2v.csv')
6 result_df_bobot_w2v_isc =
7 pd.read_csv(r'C:\Users\mit.itsupport\Downloads\archive2024\Resume2-23\rank_df_bobot_w2v_isc.csv')
8
9 result_df_tanpa_bobot_tfidf =
10 pd.read_csv(r'C:\Users\mit.itsupport\Downloads\archive2024\Resume2-23\rank_df_tanpa_bobot_tfidf.csv')
11 result_df_tanpa_bobot_w2v =
12 pd.read_csv(r'C:\Users\mit.itsupport\Downloads\archive2024\Resume2-23\rank_df_tanpa_bobot_w2v.csv')
13 result_df_tanpa_bobot_w2v_isc =
14 pd.read_csv(r'C:\Users\mit.itsupport\Downloads\archive2024\Resume2-23\rank_df_tanpa_bobot_w2v_isc.csv')

15 result_df_bobot_tfidf =
16 result_df_bobot_tfidf.drop(columns=["Link_Gdrive"])
17 result_df_bobot_w2v =
18 result_df_bobot_w2v.drop(columns=["Link_Gdrive"])
19 result_df_bobot_w2v_isc =
20 result_df_bobot_w2v_isc.drop(columns=["Link_Gdrive"])

21 result_df_tanpa_bobot_tfidf =
22 result_df_tanpa_bobot_tfidf.drop(columns=["Link_Gdrive"])

```

## Kode Program 5.12 Implementasi kode program perhitungan *Spearman Rank Correlation Coefficient* (lanjutan)

```

15     result_df_tanpa_bobot_w2v =
16     result_df_tanpa_bobot_w2v.drop(columns=["Link_Gdrive"])
17     result_df_tanpa_bobot_w2v_isc =
18     result_df_tanpa_bobot_w2v_isc.drop(columns=["Link_Gdrive"])
19
20     # Korelasi Ranking SRCC
21     def calculate_srcc(df):
22         df['d_i'] = df['Rank'] - df['Rank_Expert'] # Selisih
23         peringkat (d_i)
24
25         df['d_i_squared'] = df['d_i'] ** 2 # Kuadrat selisih
26         peringkat (d_i^2)
27
28         sum_d_i_squared = df['d_i_squared'].sum() # Total kuadrat
29         selisih peringkat ( $\sum d_i^2$ )
30
31         n = len(df)
32
33         # Hitung SRCC
34         if n < 2 or n * (n**2 - 1) == 0: # Mengatasi pembagian
35             dengan 0
36             return None
37         srcc = 1 - ((6 * sum_d_i_squared) / (n * (n**2 - 1)))
38
39         return srcc
40
41     def calculate_srcc_per_position(df):
42         results = {}
43         for position, group in df.groupby('Position'): # Hitung
44             SRCC berdasarkan 'Position'
45             srcc = calculate_srcc(group)
46             if srcc is not None:
47                 results[position] = srcc
48         return results
49
50     # Implementasi Fungsi SRCC
51     srcc_bobot_tfidf =
52     calculate_srcc_per_position(result_df_bobot_tfidf)
53     srcc_bobot_w2v =
54     calculate_srcc_per_position(result_df_bobot_w2v)
55     srcc_bobot_w2v_isc =
56     calculate_srcc_per_position(result_df_bobot_w2v_isc)
57
58     srcc_tanpa_bobot_tfidf =
59     calculate_srcc_per_position(result_df_tanpa_bobot_tfidf)
60     srcc_tanpa_bobot_w2v =
61     calculate_srcc_per_position(result_df_tanpa_bobot_w2v)
62     srcc_tanpa_bobot_w2v_isc =
63     calculate_srcc_per_position(result_df_tanpa_bobot_w2v_isc)
64
65     # Merge semua df ke satu df
66     df_bobot_tfidf = pd.DataFrame(list(srcc_bobot_tfidf.items()),
67                                   columns=['Position', 'TFIDF_Bobot'])
68     df_bobot_w2v = pd.DataFrame(list(srcc_bobot_w2v.items()),
69                                   columns=['Position', 'W2V_Bobot'])
70     df_bobot_w2v_isc =
71     pd.DataFrame(list(srcc_bobot_w2v_isc.items()),
72                  columns=['Position', 'W2V_ISC_Bobot'])
73     df_tanpa_bobot_tfidf =
74     pd.DataFrame(list(srcc_tanpa_bobot_tfidf.items()),
75                  columns=['Position', 'TFIDF_Tanpa_Bobot'])

```

## Kode Program 5.12 Implementasi kode program perhitungan *Spearman Rank Correlation Coefficient* (lanjutan)

```
57 df_tanpa_bobot_w2v =
58 pd.DataFrame(list(srcc_tanpa_bobot_w2v.items()),
59 columns=['Position', 'W2V_Tanpa_Bobot'])
60 df_tanpa_bobot_w2v_isc =
61 pd.DataFrame(list(srcc_tanpa_bobot_w2v_isc.items()),
62 columns=['Position', 'W2V_ISC_Tanpa_Bobot'])
63
64 # Jadikan 'Position' sebagai index
65 merged_df = df_bobot_tfidf.set_index('Position')
66 merged_df = merged_df.join(df_bobot_w2v.set_index('Position'),
67 how='outer')
68 merged_df =
69 merged_df.join(df_bobot_w2v_isc.set_index('Position'),
70 how='outer')
71 merged_df =
72 merged_df.join(df_tanpa_bobot_tfidf.set_index('Position'),
73 how='outer')
74 merged_df =
75 merged_df.join(df_tanpa_bobot_w2v.set_index('Position'),
76 how='outer')
77 merged_df =
78 merged_df.join(df_tanpa_bobot_w2v_isc.set_index('Position'),
79 how='outer')

# Pemberian warna untuk visualisasi
def color_srcc(val):
    if pd.isna(val): # Jika ada NaN
        return ''
    if val >= 0,6: # Kuat (hijau)
        return 'background-color: lightgreen'
    else: # Lemah (merah)
        return 'background-color: lightcoral'

styled_df =
merged_df.style.format("{:.16f}").applymap(color_srcc)
styled_df
```

Penjelasan dari Kode Program 5.12 mengenai implementasi kode program pengujian bagian perhitungan *Spearman Rank Correlation Coefficient* (SRCC), yaitu:

1. Baris 2 merupakan proses untuk menginput file dengan format CSV rank\_df\_bobot\_tfidf ke dalam DataFrame result\_df\_bobot\_tfidf menggunakan pd.read\_csv().
2. Baris 3 merupakan proses untuk menginput file dengan format CSV rank\_df\_bobot\_w2v ke dalam DataFrame result\_df\_bobot\_w2v menggunakan pd.read\_csv().
3. Baris 4 merupakan proses untuk menginput file dengan format CSV rank\_df\_bobot\_w2v\_isc ke dalam DataFrame result\_df\_bobot\_w2v\_isc menggunakan pd.read\_csv().

4. Baris 6 merupakan proses untuk menginput *file* dengan format CSV `rank_df_tanpa_bobot_tfidf` ke dalam `DataFrame result_df_tanpa_bobot_tfidf` menggunakan `pd.read_csv()`.
5. Baris 7 merupakan proses untuk menginput *file* dengan format CSV `rank_df_tanpa_bobot_w2v` ke dalam `DataFrame result_df_tanpa_bobot_w2v` menggunakan `pd.read_csv()`.
6. Baris 8 merupakan proses untuk menginput *file* dengan format CSV `rank_df_tanpa_bobot_w2v_isc` ke dalam `DataFrame result_df_tanpa_bobot_w2v_isc` menggunakan `pd.read_csv()`.
7. Baris 10-16 merupakan proses untuk menghapus kolom `Link_Gdrive` dari enam `DataFrame` yang sudah diinput.
8. Baris 18-19 merupakan proses untuk mendefinisikan fungsi `calculate_srcc` yang menerima `DataFrame df`.
9. Baris 20 merupakan proses untuk menghitung selisih antara kolom `Rank` dan `Rank_Expert` yang disimpan pada kolom baru `d_i`.
10. Baris 22 merupakan proses untuk menghitung kuadrat dari kolom `d_i` yang disimpan ke kolom baru `d_i_squared`.
11. Baris 24 merupakan proses untuk menjumlahkan semua nilai di kolom `d_i_squared` menggunakan `sum()` dan disimpan ke variabel `sum_d_i_squared`.
12. Baris 26 merupakan proses untuk menghitung jumlah baris di `DataFrame df` menggunakan `len()` dan menyimpannya ke variabel `n`.
13. Baris 28-30 merupakan proses untuk memeriksa apakah `n` kurang dari 2 atau menyebut yang merupakan rumus SRCC ( $n * (n^{**2} - 1)$ ) sama dengan 0 lalu mengembalikan `None` jika kondisi terpenuhi (`true`) untuk menghindari pembagian dengan nol.
14. Baris 31 merupakan proses untuk menghitung *Spearman Rank Correlation Coefficient* (SRCC) menggunakan persamaan  $1 - (6 * \text{sum}_d_i_squared) / (n * (n^{**2} - 1))$  dan disimpan ke variabel `srcc`.
15. Baris 33 merupakan proses untuk mengembalikan nilai `srcc` sebagai hasil dari fungsi `calculate_srcc`.
16. Baris 35-36 merupakan proses untuk mendefinisikan fungsi `calculate_srcc_per_position` dan pembuatan *dictionary* kosong `results` untuk menyimpan nilai hasil perhitungan SRCC setiap kualifikasi lowongan kerja.
17. Baris 37-40 mengelompokkan `DataFrame df` berdasarkan kolom `Position`, lalu menghitung SRCC untuk setiap kualifikasi lowongan kerja (`Position`) dengan memanggil fungsi `calculate_srcc` dan disimpan hasil-hasilnya ke *dictionary* `results` jika fungsi tersebut tidak mengembalikan `None`.

18. Baris 41 merupakan proses untuk mengembalikan isian *dictionary results* yang merupakan nilai-nilai SRCC per kualifikasi lowongan kerja sebagai hasil dari fungsi `calculate_srcc_per_position`.
19. Baris 43-50 merupakan proses untuk menghitung SRCC per kualifikasi lowongan kerja pada enam DataFrame, yaitu:
- Hasil SRCC `result_df_bobot_tfidf` disimpan ke variabel `srcc_bobot_tfidf`,
  - Hasil SRCC `result_df_bobot_w2v` disimpan ke variabel `srcc_bobot_w2v`,
  - Hasil SRCC `result_df_bobot_w2v_isc` disimpan ke variabel `srcc_bobot_w2v_isc`,
  - Hasil SRCC `result_df_tanpa_bobot_tfidf` disimpan ke variabel `srcc_tanpa_bobot_tfidf`,
  - Hasil SRCC `result_df_tanpa_bobot_w2v` disimpan ke variabel `srcc_tanpa_bobot_w2v`,
  - Hasil SRCC `result_df_tanpa_bobot_w2v_isc` disimpan ke variabel `srcc_tanpa_bobot_w2v_isc`.
20. Baris 52-58 merupakan proses untuk membuat DataFrame menggunakan `pd.DataFrame()` dari masing-masing hasil enam perhitungan SRCC per posisi yang disimpan menjadi:
- `df_bobot_tfidf` dari *dictionary srcc\_bobot\_tfidf* dengan kolom Position dan TFIDF\_Bobot,
  - `df_bobot_w2v` dari *dictionary srcc\_bobot\_w2v* dengan kolom Position dan W2V\_Bobot,
  - `df_bobot_w2v_isc` dari *dictionary srcc\_bobot\_w2v\_isc* dengan kolom Position dan W2V\_ISC\_Bobot,
  - `df_tanpa_bobot_tfidf` dari *dictionary srcc\_tanpa\_bobot\_tfidf* dengan kolom Position dan TFIDF\_Tanpa\_Bobot,
  - `df_tanpa_bobot_w2v` dari *dictionary srcc\_tanpa\_bobot\_w2v* dengan kolom Position dan W2V\_Tanpa\_Bobot,
  - `df_tanpa_bobot_w2v_isc` dari *dictionary srcc\_tanpa\_bobot\_w2v\_isc* dengan kolom Position dan W2V\_ISC\_Tanpa\_Bobot.
21. Baris 60-66 merupakan proses mengatur kolom Position sebagai *indeks* dan menggabungkan enam DataFrame untuk dijadikan satu menggunakan `df.join()` dan disimpan ke DataFrame `merged_df`.
22. Baris 68-75 merupakan proses untuk mendefinisikan fungsi `color_srcc` yang memberikan warna pada isian nilai di DataFrame berdasarkan nilai SRCC, di mana jika kosong atau NaN, maka tidak diberi warna; jika lebih dari sama dengan 0,6, maka diberi warna hijau; dan jika kurang dari 0,6, maka diberi warna merah.

23. Baris 77 merupakan proses untuk memformat DataFrame merged\_df menggunakan style.format("{:.16f}") dan menerapkan fungsi color\_src untuk pemberian warna menggunakan applymap(). Kemudian, disimpan pada DataFrame baru dengan variabel styled\_df.
24. Baris 78 merupakan proses untuk menampilkan DataFrame styled\_df.

### 5.7.2 Implementasi Kode Program Perhitungan Relevansi dan Senioritas

Dalam kode program ini, dilakukan perhitungan persentase relevansi dan senioritas per kualifikasi lowongan kerja (Position) untuk enam DataFrame dengan “Tanpa Bobot” untuk metode TF-IDF dengan *Improved Sqrt-Cosine (ISC) Similarity*, Word2Vec dengan *Cosine Similarity* (CosSim), dan Word2Vec dengan *ISC* menggunakan fungsi calculate\_relevance\_seniority\_per\_position. Tampilan hasilnya dibuat menjadi dua DataFrame agar lebih mudah dalam segi pembacaan, yakni df\_relevance untuk relevansi dan df\_seniority untuk senioritas. Kedua DataFrame tersebut divisualisasikan dengan pewarnaan berdasarkan nilai persentase, ditandai dengan warna hijau jika persentase pada suatu kualifikasi lowongan kerja (Position) bernilai di atas ambang batas dan ditandai dengan warna merah jika persentase di bawah ambang batas. Implementasi kode program tertera pada Kode Program 5.13.

#### Kode Program 5.13 Implementasi kode program perhitungan relevansi dan senioritas

```

1   # Menghitung persentase Relevansi dan Senioritas per posisi
2   def calculate_relevance_seniority_per_position(df):
3       results = {}
4       for position, group in df.groupby('Position'):
5           total_resumes = len(group)
6
7           # Hitung jumlah TRUE untuk Relevansi dan Senioritas
8           relevance_count = (group['Relevance'] == True).sum()
9           seniority_count = (group['Seniority'] == True).sum()
10
11          # Hitung persentase
12          relevance_percent = (relevance_count / total_resumes) *
13              100
14          seniority_percent = (seniority_count / total_resumes) *
15              100
16
17          # Simpan hasil
18          results[position] = {'Relevance': relevance_percent,
19                               'Seniority': seniority_percent}
20
21      return results
22
23  # Implementasi Fungsi Relevansi dan Senioritas
24  relevance_seniority_bobot_tfidf =
25  calculate_relevance_seniority_per_position(result_df_bobot_tfid
26  f)
27  relevance_seniority_bobot_w2v =
28  calculate_relevance_seniority_per_position(result_df_bobot_w2v)

```

### Kode Program 5.13 Implementasi kode program perhitungan relevansi dan senioritas (lanjutan)

```
23     relevance_seniority_bobot_w2v_isc =
24         calculate_relevance_seniority_per_position(result_df_bobot_w2v_
25             isc)
26         relevance_seniority_tanpa_bobot_tfidf =
27             calculate_relevance_seniority_per_position(result_df_tanpa_bobo
28                 t_tfidf)
29         relevance_seniority_tanpa_bobot_w2v =
30             calculate_relevance_seniority_per_position(result_df_tanpa_bobo
31                 t_w2v)
32         relevance_seniority_tanpa_bobot_w2v_isc =
33             calculate_relevance_seniority_per_position(result_df_tanpa_bobo
34                 t_w2v_isc)
35
36     # Membedakan DataFrames untuk Relevansi dan Senioritas
37     df_relevance_bobot_tfidf = pd.DataFrame(
38         [(pos, val['Relevance']) for pos, val in
39             relevance_seniority_bobot_tfidf.items()],
40         columns=['Position', 'TFIDF_Bobot']
41     )
42     df_relevance_bobot_w2v = pd.DataFrame(
43         [(pos, val['Relevance']) for pos, val in
44             relevance_seniority_bobot_w2v.items()],
45         columns=['Position', 'W2V_Bobot']
46     )
47     df_relevance_bobot_w2v_isc = pd.DataFrame(
48         [(pos, val['Relevance']) for pos, val in
49             relevance_seniority_bobot_w2v_isc.items()],
50         columns=['Position', 'W2V_ISC_Bobot']
51     )
52     df_relevance_tanpa_bobot_tfidf = pd.DataFrame(
53         [(pos, val['Relevance']) for pos, val in
54             relevance_seniority_tanpa_bobot_tfidf.items()],
55         columns=['Position', 'TFIDF_Tanpa_Bobot']
56     )
57     df_relevance_tanpa_bobot_w2v = pd.DataFrame(
58         [(pos, val['Relevance']) for pos, val in
59             relevance_seniority_tanpa_bobot_w2v.items()],
60         columns=['Position', 'W2V_Tanpa_Bobot']
61     )
62     df_relevance_tanpa_bobot_w2v_isc = pd.DataFrame(
63         [(pos, val['Relevance']) for pos, val in
64             relevance_seniority_tanpa_bobot_w2v_isc.items()],
65         columns=['Position', 'W2V_ISC_Tanpa_Bobot']
66     )
67
68     df_seniority_bobot_tfidf = pd.DataFrame(
69         [(pos, val['Seniority']) for pos, val in
70             relevance_seniority_bobot_tfidf.items()],
71         columns=['Position', 'TFIDF_Bobot']
72     )
73     df_seniority_bobot_w2v = pd.DataFrame(
74         [(pos, val['Seniority']) for pos, val in
75             relevance_seniority_bobot_w2v.items()],
76         columns=['Position', 'W2V_Bobot']
77     )
78     df_seniority_bobot_w2v_isc = pd.DataFrame(
79         [(pos, val['Seniority']) for pos, val in
80             relevance_seniority_bobot_w2v_isc.items()],
81         columns=['Position', 'W2V_ISC_Bobot']
82     )
```

### Kode Program 5.13 Implementasi kode program perhitungan relevansi dan senioritas (lanjutan)

```
66 df_seniority_tanpa_bobot_tfidf = pd.DataFrame(
67     [(pos, val['Seniority']) for pos, val in
68      relevance_seniority_tanpa_bobot_tfidf.items()],
69      columns=['Position', 'TFIDF_Tanpa_Bobot']
70 )
71 df_seniority_tanpa_bobot_w2v = pd.DataFrame(
72     [(pos, val['Seniority']) for pos, val in
73      relevance_seniority_tanpa_bobot_w2v.items()],
74      columns=['Position', 'W2V_Tanpa_Bobot']
75 )
76 df_seniority_tanpa_bobot_w2v_isc = pd.DataFrame(
77     [(pos, val['Seniority']) for pos, val in
78      relevance_seniority_tanpa_bobot_w2v_isc.items()],
79      columns=['Position', 'W2V_ISC_Tanpa_Bobot']
80 )
81 merged_relevance_df =
82 df_relevance_bobot_tfidf.set_index('Position')
83 merged_relevance_df =
84 merged_relevance_df.join(df_relevance_bobot_w2v.set_index('Position'), how='outer')
85 merged_relevance_df =
86 merged_relevance_df.join(df_relevance_bobot_w2v_isc.set_index('Position'), how='outer')
87 merged_relevance_df =
88 merged_relevance_df.join(df_relevance_tanpa_bobot_tfidf.set_index('Position'), how='outer')
89 merged_relevance_df =
90 merged_relevance_df.join(df_relevance_tanpa_bobot_w2v.set_index('Position'), how='outer')
91 merged_relevance_df =
92 merged_relevance_df.join(df_relevance_tanpa_bobot_w2v_isc.set_index('Position'), how='outer')
93
94 # Pemberian warna untuk visualisasi
95 def color_percentage(val):
96     if pd.isna(val): # Jika ada NaN
97         return ''
98     if val >= 60: # Kuat (hijau)
99         return 'background-color: lightgreen'
100    else: # Lemah (merah)
101        return 'background-color: lightcoral'
```

### Kode Program 5.13 Implementasi kode program perhitungan relevansi dan senioritas (lanjutan)

```
102     styled_relevance_df =  
103         merged_relevance_df.style.format("{:.2f}%").applymap(color_perc  
104         entage)  
105     styled_seniority_df =  
106         merged_seniority_df.style.format("{:.2f}%").applymap(color_perc  
107         entage)  
108     styled_relevance_df  
109     styled_seniority_df
```

Penjelasan dari Kode Program 5.13 mengenai implementasi kode program pengujian bagian perhitungan relevansi dan senioritas, yaitu:

1. Baris 1-3 merupakan proses untuk mendefinisikan fungsi `calculate_relevance_seniority_per_position` yang menerima DataFrame `df` dan membuat *dictionary* kosong `results` untuk menyimpan hasil perhitungan persentase relevansi dan senioritas.
2. Baris 5 merupakan proses untuk mengelompokkan DataFrame `df` berdasarkan kolom `Position` dan memulai iterasi untuk setiap kualifikasi lowongan kerja (`Position`).
3. Baris 6 merupakan proses untuk menghitung jumlah resume yang ada di dalam kualifikasi lowongan kerja menggunakan `len()` dan menyimpannya ke variabel `total_resumes`.
4. Baris 8-9 merupakan proses untuk menghitung jumlah nilai `True` di kolom `Relevance` menggunakan `(group['Relevance'] == True).sum()` dan di kolom `Seniority` menggunakan `(group['Seniority'] == True).sum()`. Kemudian, disimpan ke variabel `relevance_count` dan `seniority_count`.
5. Baris 12-14 merupakan proses untuk menghitung persentase relevansi dengan membagi `relevance_count` dengan `total_resumes` dan mengalikan dengan 100 agar hasilnya dalam bentuk persen, lalu disimpan ke variabel `relevance_percent`. Hal yang sama dilakukan pada perhitungan persentase senioritas dengan membagi `seniority_count` dengan `total_resumes` dan mengalikan dengan 100 lalu disimpan ke variabel `seniority_percent`.
6. Baris 15-18 merupakan proses untuk menambahkan nilai `position`, `relevance_percent`, dan `seniority_percent` ke *dictionary* `results`. Kemudian, dikonversi menjadi DataFrame menggunakan `pd.DataFrame()`.

7. Baris 20-26 merupakan proses untuk menghitung persentase relevansi dan senioritas per kualifikasi lowongan kerja pada enam DataFrame, yaitu:
  - a. Hasil perhitungan relevansi dan senioritas `result_df_bobot_tfidf` disimpan ke variabel `relevance_seniority_bobot_tfidf`,
  - b. Hasil perhitungan relevansi dan senioritas `result_df_bobot_w2v` disimpan ke variabel `relevance_seniority_bobot_w2v`,
  - c. Hasil perhitungan relevansi dan senioritas `result_df_bobot_w2v_isc` disimpan ke variabel `relevance_seniority_bobot_w2v_isc`,
  - d. Hasil perhitungan relevansi dan senioritas `result_df_tanpa_bobot_tfidf` disimpan ke variabel `relevance_seniority_tanpa_bobot_tfidf`,
  - e. Hasil perhitungan relevansi dan senioritas `result_df_tanpa_bobot_w2v` disimpan ke variabel `relevance_seniority_tanpa_bobot_w2v`,
  - f. Hasil perhitungan relevansi dan senioritas `result_df_tanpa_bobot_w2v_isc` disimpan ke variabel `relevance_seniority_tanpa_bobot_w2v_isc`.
8. Baris 28-52 merupakan proses untuk membuat enam DataFrame terpisah, yakni menyimpan masing-masing data relevansi dari setiap hasil perhitungan `result_df` dengan kolom `Position`, serta kolom nama pendekatan dan skenario yang digunakan.
9. Baris 54-77 merupakan proses untuk membuat enam DataFrame terpisah, yakni menyimpan masing-masing data senioritas dari setiap hasil perhitungan `result_df` dengan kolom `Position`, serta kolom nama pendekatan dan skenario yang digunakan.
10. Baris 79-84 merupakan proses mengatur kolom `Position` sebagai *indeks* dan mengabungkan enam DataFrame hasil data relevansi menggunakan `df.join()` dan disimpan ke DataFrame `merged_relevance_df`.
11. Baris 86-91 merupakan proses mengatur kolom `Position` sebagai *indeks* dan mengabungkan enam DataFrame hasil data senioritas menggunakan `df.join()` dan disimpan ke DataFrame `merged_relevance_df`.
12. Baris 93-100 merupakan proses untuk mendefinisikan fungsi `color_percentage` yang memberikan warna pada isian nilai di DataFrame berdasarkan persentase relevansi atau senioritas, di mana jika kosong atau NaN, maka tidak diberi warna; jika lebih dari sama dengan 60 maka diberi warna hijau; dan jika kurang dari 60 maka diberi warna merah.
13. Baris 102 merupakan proses untuk memformat DataFrame `relevance_df` menggunakan `style.format("{:.2f}%")` dan menerapkan fungsi `color_percentage` untuk pemberian warna menggunakan `applymap()`. Kemudian, disimpan pada DataFrame baru dengan variabel `styled_relevance_df`.

14. Baris 103 merupakan proses untuk memformat DataFrame seniority\_df menggunakan style.format("{:.2f}%") dan menerapkan fungsi color\_percentage untuk pemberian warna menggunakan applymap(). Kemudian, disimpan hasilnya ke variabel styled\_seniority\_df.
15. Baris 105 merupakan proses untuk menampilkan DataFrame styled\_relevance\_df.
16. Baris 107 merupakan proses untuk menampilkan DataFrame styled\_seniority\_df.

## BAB 6 PENGUJIAN DAN ANALISIS HASIL

Pemaparan hasil dari pengujian akan dijelaskan di bab pengujian, serta pembahasan dan analisa dari hasil pengujian tersebut sebagai bahan evaluasi.

### 6.1 Pengujian

Penelitian ini menggunakan dua pendekatan perhitungan representasi teks, yakni TF-IDF dan Word2Vec. Untuk pendekatan perhitungan similaritasnya dibuat menjadi tiga kombinasi pendekatan, yakni TF-IDF dengan *Improved Sqrt-Cosine (ISC) Similarity*, Word2Vec dengan *Cosine Similarity* (CosSim), dan Word2Vec dengan ISC. Masing-masing pendekatan diuji dengan dua skenario. Skenario pertama adalah lima resume dengan skor similaritas terbesar jika tanpa menggunakan pembobotan per *section* dari ahli dan skenario kedua adalah lima resume dengan skor similaritas terbesar jika menggunakan pembobotan per *section* dari ahli. Dalam masing-masing kombinasi pendekatan dan skenario terdapat tiga parameter penilaian untuk ahli memberikan *ground truth* berdasarkan hasil keluaran implementasi dan skenario, yakni peringkat (*rank*), relevansi, dan senioritas, dengan fokus utama pada parameter peringkat. Parameter relevansi dan senioritas berperan sebagai pendukung untuk memperkaya evaluasi. Implementasi metode atau pendekatan menghasilkan skor similaritas antara setiap resume dengan masing-masing kualifikasi lowongan kerja. Untuk setiap kualifikasi lowongan kerja, dipilih lima resume dengan skor similaritas tertinggi. Kelima resume tersebut selanjutnya dievaluasi oleh ahli berdasarkan tiga parameter, yakni: (1) urutan peringkat yang dianggap paling sesuai (*ground truth*), (2) relevansi isi resume terhadap kualifikasi lowongan kerja, dan (3) kesesuaian level posisi resume dengan level posisi yang diminta dalam kualifikasi.

Lima peringkat teratas hasil keluaran dari implementasi metode dipindahkan ke dalam *spreadsheet* saat disajikan kepada ahli untuk memudahkan proses evaluasi. Cuplikan *template* penyajian tersebut tertera pada Gambar 6.1 dan Gambar 6.2.

Rank	CV_ID	Position	Similarity_Score	Link_Owner			Rank_Expert	Relevance	Seniority
				E15561087	E15561088.pdf	1			
1	30863814	Business Development Executive	0.1528144	E15561088	E15561088.pdf	2	✓	✓	✓
2	18311419	Business Development Executive	0.17471594	E15561089	E15561089.pdf	3	✓	✓	✓
3	18311419	Business Development Executive	0.17471594	E15561090	E15561090.pdf	4	✓	✓	✓
4	15539523	Business Development Executive	0.168656137	E15561092	E15561092.pdf	5	✓	✓	✓
5	47075153	Business Development Executive	0.164320994	E15561094	E15561094.pdf	6	✓	✓	✓

Gambar 6.1 Cuplikan *template spreadsheet* evaluasi ahli

Gambar 6.2 Cuplikan template spreadsheet evaluasi ahli

Hasil evaluasi dari ahli dihimpun dan dirapikan agar bisa dijadikan DataFrame untuk perhitungan ketiga parameter penilaian seperti pada Gambar 6.3

Gambar 6.3 Cuplikan spreadsheet hasil evaluasi ahli

Parameter peringkat (*rank*) dievaluasi hasil pengujinya dengan menghitung *Spearman Rank Correlation Coefficient* (SRCC) untuk menunjukkan korelasi antara peringkat yang dihasilkan implementasi metode dengan peringkat yang dibenarkan (*ground truth*) oleh ahli. Semakin tinggi nilai korelasi, maka semakin baik metode tersebut mengurutkan similaritas yang sesuai dengan pandangan ahli (manusia). Pada penelitian ini, dilakukan deskripsi statistik melalui SPSS untuk mengetahui persentil dari keseluruhan nilai korelasi seperti tertera pada Gambar 6.4 dan diketahui persentil ke-75 adalah korelasi positif 0,6. Nilai 0,6 ini menunjukkan bahwa 75% dari data lainnya memiliki nilai yang lebih rendah. Oleh karena itu, nilai  $\geq 0,6$  ditentukan sebagai ambang batas parameter korelasi.

## **Statistics**

SRCC

N	Valid	144
	Missing	0
Minimum		-1.00
Maximum		1.00
Percentiles	25	-.3000
	50	.2000
	75	.6000

**Gambar 6.4 Hasil *descriptive statistics* SPSS**

Sebagai parameter pendukung, relevansi dan senioritas pada penelitian ini dianggap baik jika setidaknya tiga dari lima resume memenuhi deskripsi kualifikasi lowongan kerja berdasarkan evaluasi ahli. Sehingga, nilai persentase  $\geq 60\%$  ditentukan sebagai ambang batas parameter relevansi dan senioritas.

Dari 24 kualifikasi lowongan kerja pada penelitian ini, SRCC dengan nilai  $\geq 0,6$  dianggap kuat (*strong*) yang ditandai dengan warna hijau. Sedangkan SRCC dengan nilai  $< 0,6$  dianggap lemah (*weak*) yang ditandai dengan warna merah. Gambar 6.5 merupakan visualisasi dari nilai korelasi setiap kombinasi pendekatan berdasarkan posisi kualifikasi lowongan kerja.

Position	TFIDF_Bobot	W2V_Bobot	W2V_ISC_Bobot	TFIDF_Tanpa_Bobot	W2V_Tanpa_Bobot	W2V_ISC_Tanpa_Bobot
Business Development Executive	0.8000000000000000	0.3000000000000000	0.3000000000000000	0.1000000000000000	-0.6000000000000001	-0.2000000000000000
CLUB GENERAL MANAGER	-0.3000000000000000	0.1000000000000000	0.1000000000000000	-0.7000000000000000	-1.0000000000000000	0.0000000000000000
Construction Supervisor	0.6000000000000000	-0.3000000000000000	-0.5000000000000000	0.9000000000000000	0.1000000000000000	0.1000000000000000
Creative Director / Manager	0.0000000000000000	0.2000000000000000	-0.8000000000000000	-0.3000000000000000	0.5000000000000000	0.2000000000000000
Digital Banking Officer	-0.3999999999999999	0.1000000000000000	-0.3000000000000000	-0.6000000000000001	0.3000000000000000	-0.3000000000000000
Digital and Social Media Executive	-0.5000000000000000	-0.8999999999999999	0.6000000000000000	-0.2000000000000000	0.1000000000000000	0.8000000000000000
Executive Chef	-0.2000000000000000	-0.7000000000000000	-0.3000000000000000	0.5000000000000000	-0.8000000000000000	-0.7000000000000000
Finance Executive / Accountant	-1.0000000000000000	0.6000000000000000	0.2000000000000000	-0.8999999999999999	0.5000000000000000	0.7000000000000000
Finance Officer ( Jr/Sr.)	0.2000000000000000	0.7000000000000000	0.9000000000000000	0.7000000000000000	0.7000000000000000	-0.1000000000000001
Financial Consolidation Consultant	-0.1000000000000001	1.0000000000000000	0.4000000000000000	-0.1000000000000001	0.7000000000000000	0.3000000000000000
Graphics Designer	0.7000000000000000	0.9000000000000000	0.5000000000000000	-0.3000000000000000	0.3000000000000000	0.5000000000000000
HR Specialist	0.8000000000000000	0.7000000000000000	0.1000000000000000	0.7000000000000000	0.2000000000000000	0.7000000000000000
INFORMATION & TECHNOLOGY STAFF	0.3000000000000000	-0.5000000000000000	0.1000000000000000	0.3000000000000000	-0.3999999999999999	0.6000000000000000
Junior Associate Lawyer	0.0000000000000000	0.3000000000000000	-0.7000000000000000	0.2000000000000000	-0.6000000000000001	1.0000000000000000
Junior Designer for Apparel	0.5000000000000000	0.6000000000000000	-0.3000000000000000	0.9000000000000000	-0.1000000000000001	-0.2000000000000000
Manager Aviation Safety, Quality and Security	0.9000000000000000	0.9000000000000000	-0.1000000000000001	-0.7000000000000000	0.9000000000000000	0.4000000000000000
Medical Doctor	0.4000000000000000	0.2000000000000000	0.1000000000000000	0.1000000000000000	0.6000000000000000	0.3000000000000000
Production Engineering	0.5000000000000000	-0.7000000000000000	0.2500000000000000	0.1000000000000000	-0.2000000000000000	0.5000000000000000
Public Relations Officer	0.3000000000000000	0.7000000000000000	0.7000000000000000	-0.5000000000000000	0.7000000000000000	0.7000000000000000
Quality Control Supervisor - Corn Commodity	-0.5000000000000000	0.1000000000000000	0.5000000000000000	0.3000000000000000	-0.3000000000000000	0.5000000000000000
Regional Sales Manager	0.1000000000000000	0.3000000000000000	0.2000000000000000	-0.6000000000000001	0.7000000000000000	-0.5000000000000000
Spare part Admin	-0.6000000000000001	-0.3999999999999999	-0.3000000000000000	-0.6000000000000001	-0.5000000000000000	0.9000000000000000
Teachers	-0.7000000000000000	-0.5000000000000000	-0.5000000000000000	0.8000000000000000	-0.5000000000000000	-0.7000000000000000
Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0.7000000000000000	0.7000000000000000	0.7000000000000000	0.0000000000000000	0.6000000000000000	0.7000000000000000

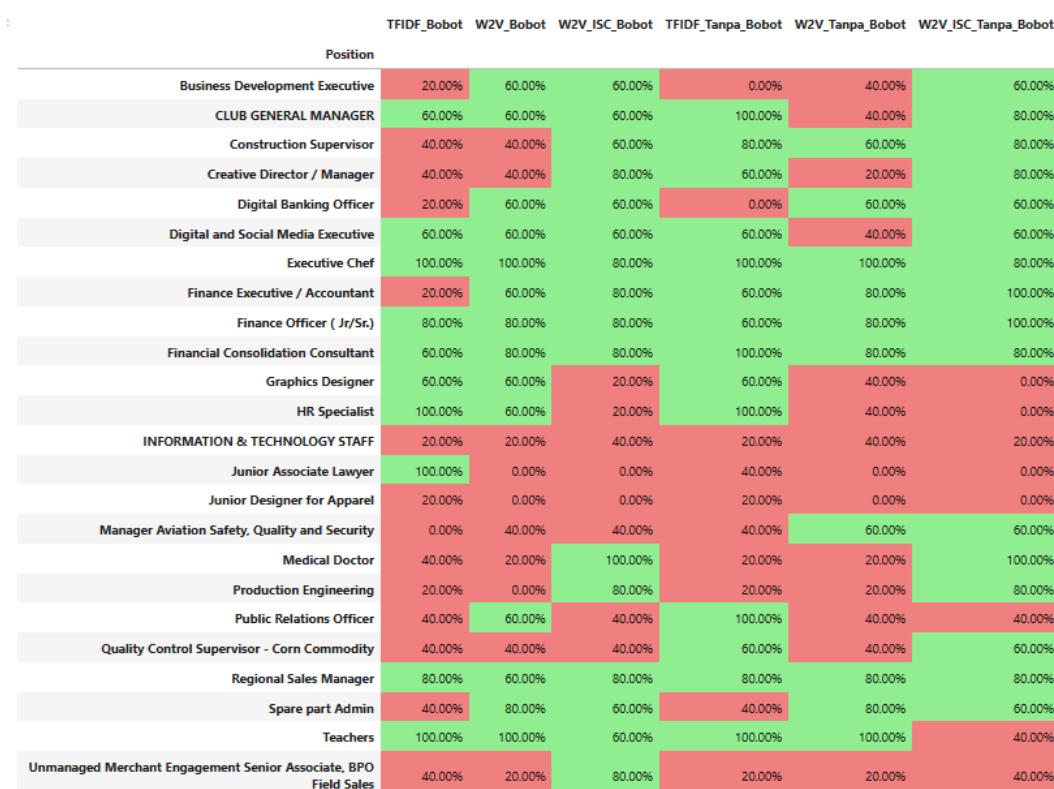
Gambar 6.5 Visualisasi nilai korelasi

Parameter relevansi (*relevance*) dievaluasi hasil pengujinya dengan menghitung persentasenya untuk menunjukkan seberapa relevan resume-resume yang menjadi keluaran implementasi metode dengan kualifikasi lowongan kerja. Semakin tinggi persentasenya menunjukkan performa yang lebih baik dalam konteks relevansi. Dari 24 kualifikasi lowongan kerja pada penelitian ini, persentase dengan nilai  $\geq 60\%$  dianggap memiliki kesesuaian yang tinggi (*high*) dan ditandai dengan warna hijau. Sedangkan persentase dengan nilai  $< 60\%$  dianggap memiliki kesesuaian yang kurang (*low*) dan ditandai dengan warna merah. Gambar 6.6 merupakan visualisasi dari nilai persentase relevansi setiap kombinasi pendekatan berdasarkan posisi kualifikasi lowongan kerja.

Position	TFIDF_Bobot	W2V_Bobot	W2V_ISC_Bobot	TFIDF_Tanpa_Bobot	W2V_Tanpa_Bobot	W2V_ISC_Tanpa_Bobot
Business Development Executive	80.00%	80.00%	80.00%	100.00%	100.00%	80.00%
CLUB GENERAL MANAGER	60.00%	40.00%	40.00%	100.00%	20.00%	20.00%
Construction Supervisor	80.00%	60.00%	60.00%	80.00%	60.00%	80.00%
Creative Director / Manager	40.00%	0.00%	20.00%	60.00%	20.00%	20.00%
Digital Banking Officer	60.00%	40.00%	0.00%	0.00%	0.00%	0.00%
Digital and Social Media Executive	80.00%	40.00%	60.00%	60.00%	40.00%	80.00%
Executive Chef	100.00%	100.00%	60.00%	100.00%	100.00%	80.00%
Finance Executive / Accountant	100.00%	80.00%	60.00%	100.00%	60.00%	40.00%
Finance Officer ( Jr/Sr.)	100.00%	80.00%	20.00%	100.00%	60.00%	40.00%
Financial Consolidation Consultant	80.00%	20.00%	20.00%	100.00%	20.00%	60.00%
Graphics Designer	80.00%	80.00%	100.00%	80.00%	60.00%	100.00%
HR Specialist	40.00%	20.00%	60.00%	40.00%	20.00%	40.00%
INFORMATION & TECHNOLOGY STAFF	40.00%	80.00%	20.00%	60.00%	60.00%	60.00%
Junior Associate Lawyer	40.00%	20.00%	0.00%	40.00%	0.00%	0.00%
Junior Designer for Apparel	40.00%	0.00%	0.00%	60.00%	20.00%	20.00%
Manager Aviation Safety, Quality and Security	20.00%	20.00%	40.00%	60.00%	40.00%	60.00%
Medical Doctor	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Production Engineering	60.00%	20.00%	20.00%	40.00%	20.00%	20.00%
Public Relations Officer	60.00%	60.00%	80.00%	60.00%	80.00%	60.00%
Quality Control Supervisor - Corn Commodity	0.00%	0.00%	0.00%	0.00%	0.00%	0.00%
Regional Sales Manager	0.00%	40.00%	60.00%	40.00%	40.00%	100.00%
Spare part Admin	40.00%	0.00%	20.00%	0.00%	0.00%	0.00%
Teachers	100.00%	80.00%	80.00%	80.00%	80.00%	100.00%
Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	40.00%	60.00%	80.00%	40.00%	40.00%	40.00%

Gambar 6.6 Visualisasi persentase relevansi

Parameter senioritas (*seniority*) dievaluasi hasil pengujinya dengan menghitung persentasenya untuk menunjukkan seberapa sesuai level posisi yang tercantum di resume-resume yang menjadi keluaran implementasi metode dengan level posisi yang dibutuhkan pada kualifikasi lowongan kerja. Semakin tinggi persentasenya menunjukkan tingkat kesesuaian yang lebih baik. Dari 24 kualifikasi lowongan kerja pada penelitian ini, persentase dengan nilai  $\geq 60\%$  dianggap memiliki kesesuaian level posisi yang tinggi (*high*) dan ditandai dengan warna hijau. Sedangkan persentase dengan nilai  $< 60\%$  dianggap memiliki kesesuaian level posisi yang rendah (*low*) dan ditandai dengan warna merah. Gambar 6.7 merupakan visualisasi dari nilai persentase senioritas setiap kombinasi pendekatan berdasarkan posisi kualifikasi lowongan kerja.



**Gambar 6.7 Visualisasi persentase senioritas**

## 6.2 Analisis Hasil

Metode *Improved Sqrt-Cosine* (ISC) *Similarity* pada penelitian ini digunakan untuk memeringkat lima resume berdasarkan skor similaritas tertinggi untuk masing-masing 24 kualifikasi lowongan kerja. Perhitungan similaritas ISC memungkinkan implementasi pada representasi teks TF-IDF yang berbasis frekuensi. Namun, untuk yang berbasis semantik menggunakan Word2Vec, vektor harus diambil nilai absolutnya karena vektor hasil Word2Vec dapat mengandung bilangan negatif. Sedangkan, rumus ISC melibatkan operasi akar kuadrat yang tidak dapat diterapkan langsung pada bilangan negatif. Oleh karena itu, Kombinasi pendekatan Word2Vec dengan *Cosine Similarity* juga digunakan untuk mempertahankan makna semantik asli dari vektor Word2Vec tanpa modifikasi nilai absolut.

Performa masing-masing pendekatan dievaluasi melalui tiga parameter, yakni *Spearman Rank Correlation Coefficient* (SRCC) untuk mengukur korelasi dengan pemeringkatan dari hasil evaluasi ahli, relevansi untuk menilai kesesuaian resume dengan kualifikasi lowongan kerja, dan senioritas untuk mengevaluasi kesesuaian level posisi berdasarkan kata-kata kunci. Pengujian dilakukan dalam dua skenario, yakni “Tanpa Bobot” dan “Dengan Bobot”, di mana bobot per *section* (misalnya, *Summary* 5%, *Experience* 20%, *Certification* 15%) mencerminkan prioritas penilaian seorang rekruter.

Berdasarkan hasil pengujian dengan skenario “Tanpa Bobot”, jumlah kualifikasi lowongan kerja yang memenuhi ambang batas ( $SRCC \geq 0,6$ ; relevansi  $\geq 60\%$ ; senioritas  $\geq 60\%$ ; ditandai hijau pada visualisasi) dan yang tidak memenuhi ambang batas (ditandai merah) pada Tabel 6.1 dan Tabel 6.2.

**Tabel 6.1 Hasil pengujian berwarna hijau skenario tanpa bobot section**

Keterangan	TF-IDF + ISC	Word2Vec + CosSim	Word2Vec + ISC
Jumlah yang nilai SRCC $\geq 0,6$	5	7	8
Jumlah yang persentase Relevansi $\geq 60\%$	15	9	11
Jumlah yang persentase Senioritas $\geq 60\%$	14	10	16

**Tabel 6.2 Hasil pengujian berwarna merah skenario tanpa bobot *section***

Keterangan	TF-IDF + ISC	Word2Vec + CosSim	Word2Vec + ISC
Jumlah yang nilai SRCC < 0,6	19	17	16
Jumlah yang persentase Relevansi < 60%	9	15	13
Jumlah yang persentase SenioritasS < 60%	10	14	8

Pada parameter korelasi, pendekatan Word2Vec dengan ISC unggul menempati urutan tertinggi berjumlah 8 kualifikasi lowongan kerja yang memenuhi ambang batas, diikuti oleh Word2Vec dengan CosSim berjumlah 7, dan TF-IDF dengan ISC di urutan terakhir berjumlah 5. Ini menunjukkan hasil pemeringkatan yang dihasilkan implementasi metode Word2Vec dengan ISC lebih dekat dengan penilaian ahli.

Pada parameter relevansi, pendekatan TF-IDF dengan ISC unggul dengan 15 kualifikasi lowongan kerja, disusul pendekatan Word2Vec dengan ISC berjumlah 11, dan Word2Vec dengan CosSim yang berada di urutan terakhir berjumlah 9. Ini menunjukkan isi dari lima resume keluaran implementasi metode TF-IDF dengan ISC lebih banyak yang sesuai dengan deskripsi kualifikasi lowongan kerja.

Sementara itu, dalam hal senioritas, pendekatan Word2Vec dengan ISC kembali unggul di urutan pertama dengan 16 kualifikasi lowongan kerja, diikuti oleh pendekatan TF-IDF dengan ISC berjumlah 14, dan Word2Vec dengan CosSim berjumlah 10. Ini menunjukkan isi dari lima resume keluaran implementasi Word2Vec dengan ISC lebih banyak yang level senioritasnya atau level posisinya yang sesuai dengan deskripsi kualifikasi lowongan kerja.

Berdasarkan hasil pengujian dengan skenario “Dengan Bobot”, jumlah kualifikasi lowongan kerja yang memenuhi ambang batas ( $SRCC \geq 0,6$ ; relevansi  $\geq 60\%$ ; senioritas  $\geq 60\%$ ; ditandai hijau pada visualisasi) dan yang tidak memenuhi ambang batas (ditandai merah).

**Tabel 6.3 Hasil pengujian berwarna hijau skenario dengan bobot section**

Keterangan	TF-IDF + ISC	Word2Vec + CosSim	Word2Vec + ISC
Jumlah yang nilai $SRCC \geq 0,6$	6	9	4
Jumlah yang persentase Relevansi $\geq 60\%$	13	10	11
Jumlah yang persentase Senioritas $\geq 60\%$	10	14	16

**Tabel 6.4 Hasil pengujian berwarna merah skenario dengan bobot section**

Keterangan	TF-IDF + ISC	Word2Vec + CosSim	Word2Vec + ISC
Jumlah yang nilai $SRCC < 0,6$	18	15	20
Jumlah yang persentase Relevansi $< 60\%$	11	14	13
Jumlah yang persentase Senioritas $< 60\%$	14	10	8

Pada parameter korelasi, pendekatan Word2Vec dengan CosSim unggul menempati urutan tertinggi dengan jumlah 9 kualifikasi, diikuti oleh TF-IDF dengan ISC berjumlah 6, dan Word2Vec dengan ISC di urutan terakhir berjumlah 4. Ini menunjukkan hasil pemeringkatan yang dihasilkan implementasi metode Word2Vec dengan CosSim lebih dekat dengan penilaian ahli.

Pada parameter relevansi, pendekatan TF-IDF dengan ISC unggul dengan jumlah 13 kualifikasi, disusul pendekatan Word2Vec dengan ISC berjumlah 11, dan Word2Vec dengan CosSim yang berada di urutan terakhir berjumlah 10. Ini menunjukkan isi dari lima resume keluaran implementasi metode TF-IDF dengan ISC lebih banyak yang sesuai dengan deskripsi kualifikasi lowongan kerja.

Sementara itu, dalam hal senioritas, pendekatan Word2Vec dengan ISC unggul di urutan pertama dengan jumlah 16 kualifikasi, diikuti oleh pendekatan Word2Vec dengan CosSim berjumlah 14, dan TF-IDF dengan ISC berjumlah 10. Ini menunjukkan isi dari lima resume keluaran implementasi metode Word2Vec dengan ISC lebih banyak yang sesuai dengan deskripsi kualifikasi lowongan kerja.

Secara jumlah kualifikasi lowongan kerja yang memenuhi dan yang tidak memenuhi ambang batas pada Tabel 6.1 hingga Tabel 6.4, dihitung menggunakan *weighted scoring*, di mana jumlah yang di atas ambang batas diberikan poin +2. Sedangkan, jumlah yang di bawah ambang batas diberikan poin -1. Hasil perhitungan *weighted score* untuk setiap pendekatan dan skenario tertera pada Tabel 6.5.

**Tabel 6.5 Weighted score keseluruhan pendekatan dan skenario**

Pendekatan	Korelasi	Relevansi	Senioritas	<i>Weighted Score</i>
TF-IDF + ISC Tanpa Bobot	$4 \times (2)$ $+ 20 \times (-1)$ $= (-12)$	$14 \times (2)$ $+ 10 \times (-1)$ $= 18$	$13 \times (2)$ $+ 11 \times (-1)$ $= 15$	30
Word2Vec + CosSim Tanpa Bobot	$7 \times (2)$ $+ 17 \times (-1)$ $= (-3)$	$9 \times (2)$ $+ 15 \times (-1)$ $= 3$	$10 \times (2)$ $+ 14 \times (-1)$ $= 6$	6
Word2Vec + ISC Tanpa Bobot	$8 \times (2)$ $+ 16 \times (-1)$ $= 0$	$11 \times (2)$ $+ 13 \times (-1)$ $= 9$	$16 \times (2)$ $+ 8 \times (-1)$ $= 24$	33
TF-IDF + ISC Dengan Bobot	$5 \times (2)$ $+ 19 \times (-1)$ $= -9$	$12 \times (2)$ $+ 12 \times (-1)$ $= 12$	$10 \times (2)$ $+ 14 \times (-1)$ $= 6$	15
Word2Vec + CosSim Dengan Bobot	$9 \times (2)$ $+ 15 \times (-1)$ $= 3$	$10 \times (2)$ $+ 14 \times (-1)$ $= 6$	$14 \times (2)$ $+ 10 \times (-1)$ $= 18$	27
Word2Vec + ISC Dengan Bobot	$4 \times (2)$ $+ 20 \times (-1)$ $= (-12)$	$11 \times (2)$ $+ 13 \times (-1)$ $= 9$	$16 \times (2)$ $+ 8 \times (-1)$ $= 24$	21

Dari Tabel 6.5, diurutkan berdasarkan *weighted score* tertinggi untuk masing-masing skenario seperti pada Tabel 6.6.

**Tabel 6.6 Urutan pendekatan berdasarkan weighted score tertinggi**

Tanpa Bobot	Dengan Bobot
Word2Vec + ISC (33)	Word2Vec + CosSim (27)
TF-IDF + ISC (30)	Word2Vec + ISC (21)
Word2Vec + CosSim (6)	TF-IDF + ISC (15)

Secara rata-rata, dihitung untuk setiap parameter pada masing-masing kombinasi pendekatan dan skenario yang tertera hasilnya pada Tabel 6.7.

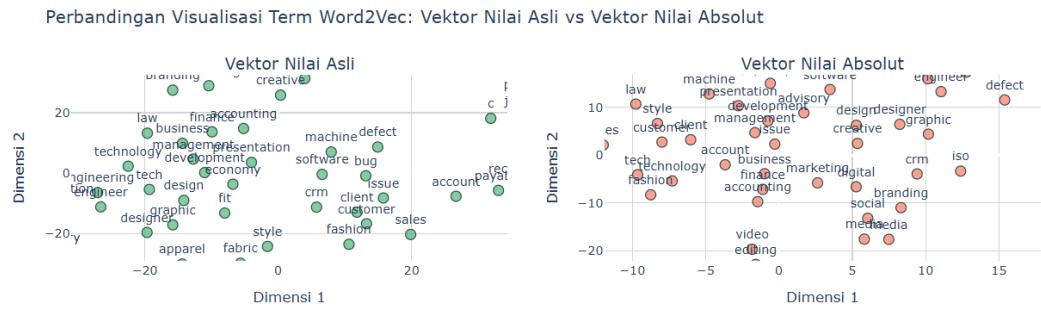
**Tabel 6.7 Perhitungan rata-rata parameter setiap pendekatan dan skenario**

Pendekatan	Korelasi	Relevansi	Senioritas
TF-IDF + ISC Tanpa Bobot	0,004348	58,33%	55,83%
Word2Vec + CosSim Tanpa Bobot	0,079167	39,17%	49,17%
Word2Vec + ISC Tanpa Bobot	0,269565	45,83%	56,67%
TF-IDF + ISC Dengan Bobot	0,113636	55,83%	50,0%
Word2Vec + CosSim Dengan Bobot	0,183333	42,5%	50,0%
Word2Vec + ISC Dengan Bobot	0,077083	40,83%	56,67%

Terlihat bahwa Word2Vec dengan ISC unggul pada skenario “Tanpa Bobot” dengan nilai korelasi rata-rata 0,269565. Namun, nilai tersebut tergolong sebagai korelasi yang lemah. Di skenario “Dengan Bobot”, Word2Vec dengan CosSim unggul dengan nilai korelasi rata-rata 0,183333 yang tergolong sebagai korelasi sangat lemah.

Dalam skenario “Tanpa Bobot” maupun “Dengan Bobot”, pendekatan ISC dengan representasi teks TF-IDF menunjukkan performa terbaik pada parameter relevansi, mengungguli Word2Vec dengan ISC dan Word2Vec dengan CosSim. Hal ini menunjukkan bahwa ISC dengan TF-IDF lebih baik dalam mengidentifikasi isi resume yang relevan dengan kualifikasi lowongan kerja berdasarkan kesesuaian kualifikasi lowongan kerja. Keunggulan TF-IDF dengan ISC pada relevansi disebabkan oleh kemampuan TF-IDF dalam memberi bobot lebih tinggi pada kata-kata penting yang jarang muncul, tetapi relevan dengan kualifikasi lowongan kerja, seperti istilah-istilah teknis.

Pendekatan Word2Vec dengan ISC menunjukkan keterbatasan akibat distorsi semantik dari penyesuaian nilai absolut, seperti terlihat dari visualisasi yang menunjukkan pergeseran posisi kata pada Gambar 6.8.



**Gambar 6.8 Visualisasi pergeseran posisi term Word2Vec vektor nilai asli dengan vektor nilai absolut**

Lebih rincinya, dilakukan perhitungan similaritas antar *term* menggunakan CosSim dengan hasil tertera pada Tabel 6.8.

**Tabel 6.8 Perhitungan similaritas antar term Word2Vec vektor nilai asli dengan vektor nilai absolut**

Term 1 versus Term 2	Similaritas (Nilai Asli)	Similaritas (Nilai Absolut)	Perubahan
<i>technology</i> versus <i>engineering</i>	0,5711	0,7413	0,1701
<i>technology</i> versus <i>fashion</i>	0,2471	0,7262	0,4791
<i>design</i> versus <i>designer</i>	0,6348	0,8274	0,1927
<i>software</i> versus <i>java</i>	0,4153	0,6950	0,2797
<i>marketing</i> versus <i>sales</i>	0,4816	0,6899	0,2083
<i>university</i> versus <i>cook</i>	0,0751	0,6303	0,5552
<i>pastry</i> versus <i>cook</i>	0,6423	0,7609	0,1185
<i>pastry</i> versus <i>economy</i>	0,2151	0,7557	0,5406
<i>bake</i> versus <i>diploma</i>	0,2106	0,6603	0,4497
<i>law</i> versus <i>style</i>	0,0141	0,7308	0,7167

Tabel 6.8 menunjukkan peningkatan atau penurunan nilai similaritas antar *term*, misalnya antara “*technology*” dengan “*fashion*” seharusnya kedua *term* ini secara semantik berjauhan dan dibuktikan dengan similaritasnya yang bernilai 0,2471. Namun, setelah nilai vektor dibuat absolut, nilai similaritasnya menjadi 0,7262 yang menyatakan kedua *term* ini berdekatan secara semantik, perubahannya cukup drastis sebanyak 0,4791. Kemudian, ada juga *term* yang sudah berdekatan secara semantik menjadi semakin dekat setelah nilainya dibuat absolut, seperti “*design*” dan “*designer*” dari similaritasnya bernilai 0,6348 menjadi 0,8274 yang menyatakan posisi kedua *term* ini menjadi semakin dekat. Visualisasi dari beberapa pasangan *term* yang dibandingkan pada Tabel 6.5 diilustrasikan pada Gambar 6.9.

Perbandingan Visualisasi Term Word2Vec: Vektor Nilai Asli vs Vektor Nilai Absolut

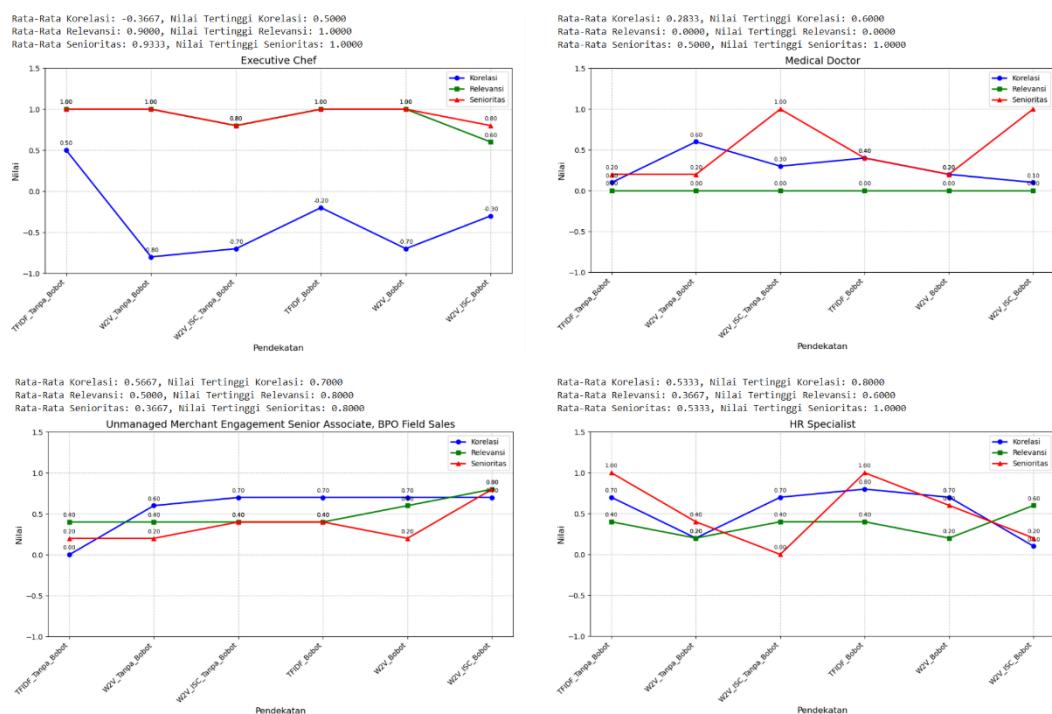


**Gambar 6.9 Visualisasi pergeseran posisi term Word2Vec vektor nilai asli dengan vektor nilai absolut**

Oleh karena itu, Word2Vec dengan ISC unggul dalam skenario “Tanpa Bobot” disebabkan oleh kemampuan representasi semantik Word2Vec yang masih cukup mampu menangkap hubungan antar kata kunci yang relevan, meskipun distorsi terjadi.

Keunggulan Word2Vec dengan CosSim dalam skenario “Dengan Bobot” menunjukkan bahwa pendekatan ini paling selaras dengan pandangan manusia karena CosSim menjaga hubungan semantik asli vektor Word2Vec tanpa distorsi akibat penyesuaian nilai absolut, dan bobot per *section* memungkinkan penyesuaian skor sesuai prioritas rekruter, seperti penekanan pada pengalaman kerja. Namun, perlu dicatat bahwa bobot per *section* di penelitian ini hanya didasarkan pada satu ahli yang mungkin tidak mencerminkan variasi preferensi rekruter dari berbagai industri atau budaya instansi terkait. Pemberian bobot dengan melibatkan lebih banyak ahli akan meningkatkan hasil secara *general*.

Melihat lebih dekat setiap kualifikasi lowongan kerja, ketiga parameter dibuat grafik garis, dengan parameter korelasi diwarnai biru, parameter relevansi diwarnai hijau, dan parameter senioritas diwarnai merah. Sumbu Y menunjukkan nilai, sehingga nilai persentase relevansi dan senioritas dijadikan bilangan desimal. Sumbu X menunjukkan pendekatan dan skenario yang digunakan. Grafik untuk beberapa kualifikasi lowongan kerja tertera pada Gambar 6.10.



**Gambar 6.10 Grafik garis tiga parameter setiap kualifikasi lowongan kerja**

Nilai korelasi pada lowongan kerja posisi “Executive Chef” memiliki nilai tertinggi 0,5 dan terlihat jauh di bawah nilai relevansi dan senioritasnya. Nilai korelasi tertinggi tersebut tergolong sedang. Sementara itu, nilai korelasi lainnya bernilai negatif dan tergolong sebagai korelasi sangat lemah. Hal ini menunjukkan walaupun korelasi antara hasil implementasi dengan hasil evaluasi ahli tergolong lemah, keluaran resume yang dihasilkan untuk posisi ini tetap sesuai dengan deskripsi kualifikasi lowongan kerja berdasarkan kesesuaian kata kunci atau istilah-istilah tertentu yang dimiliki resume maupun pada kebutuhan posisi.

Nilai korelasi pada posisi “Medical Doctor” cukup stabil dengan rata-rata 0,2833 di antara berbagai implementasi pendekatan dan skenario, meskipun tergolong sebagai korelasi lemah. Namun, nilai relevansi menunjukkan 0 yang menandakan tidak ada resume yang isiannya sesuai dengan kebutuhan posisi tersebut. Hal ini menunjukkan perlu adanya penyesuaian *dataset* resume dengan menambahkan resume yang mencakup beragam spesialisasi dalam industri *healthcare* atau sebaliknya dengan memperkaya *dataset* kualifikasi lowongan kerja di industri *healthcare* yang lebih variatif spesialisasinya.

Nilai korelasi pada posisi “Unmanaged Merchant Engagement Senior Associate, BPO Field Sales” tergolong korelasi sedang ke korelasi kuat berdasarkan nilai rata-rata 0,5667 dan nilai tertinggi 0,7. Namun, untuk nilai senioritasnya memiliki rata-rata 0,3667 atau sekitar 36%. Hal ini menunjukkan kurang mampunya hasil implementasi dalam mengenali kesuaian level senioritas dengan yang dibutuhkan pada posisi tersebut. Sehingga, perlu adanya penyesuaian *dataset* resume maupun kualifikasi lowongan kerja yang lebih beragam level senioritasnya atau pengelompokan terpisah berdasarkan level senioritas berdasarkan *years of experience* di dalam resume.

Nilai korelasi pada posisi “HR Specialist” tergolong korelasi sedang ke korelasi kuat berdasarkan nilai rata-rata 0,5333 dan nilai tertinggi 0,8. Namun, untuk nilai relevansinya memiliki rata-rata 0,3667 atau sekitar 36%. Hal ini menunjukkan kurang mampunya hasil implementasi dalam mengenali kesesuaian kata kunci atau istilah dengan yang dibutuhkan pada posisi tersebut. Sehingga, perlu adanya penyesuaian *dataset* resume maupun kualifikasi lowongan kerja yang memperbanyak kata-kata penting seperti kata kunci atau istilah teknis sesuai industri atau bidang minat.

## BAB 7 PENUTUP

Pada bab penutup, bagian kesimpulan merangkum hasil penelitian untuk menjawab rumusan masalah. Sedangkan, bagian saran memberikan masukan untuk perbaikan dan pengembangan penelitian selanjutnya.

### 7.1 Kesimpulan

Berikut merupakan kesimpulan dari penelitian ini untuk menjawab rumusan masalah.

1. Hasil pemeringkatan lima resume menggunakan *Improved Sqrt-Cosine (ISC) Similarity* dalam mengkalkulasikan similaritas teks resume dengan kualifikasi lowongan kerja menunjukkan bahwa pendekatan representasi teks menggunakan Word2Vec lebih unggul dibandingkan dengan TF-IDF untuk skenario “Tanpa Bobot”. Meskipun terdapat distorsi akibat penyesuaian nilai absolut, mengingat adanya *term* yang jarak posisinya sudah dekat dan menjadi semakin dekat, Word2Vec cukup mampu menangkap hubungan semantik kata kunci. Pada skenario “Dengan Bobot”, pendekatan perhitungan similaritas menggunakan *Cosine Similarity (CosSim)* dengan representasi teks Word2Vec lebih unggul. Namun, dalam penggunaan ISC, representasi teks Word2Vec tetap unggul dibandingkan TF-IDF. Keunggulan Word2Vec dengan CosSim menunjukkan bahwa pendekatan ini paling selaras dengan pandangan manusia karena CosSim menjaga hubungan semantik asli vektor Word2Vec tanpa distorsi akibat penyesuaian nilai absolut dan memungkinkan penyesuaian bobot per *section* sesuai prioritas seorang ahli dalam penelitian ini.
2. Korelasi peringkat antara hasil pemeringkatan dari ISC dengan penilaian ahli terhadap kesesuaian kualifikasi lowongan kerja menunjukkan keunggulan yang bervariasi berdasarkan representasi teks dan skenario, di mana pada skenario “Tanpa Bobot”, Word2Vec dengan ISC lebih unggul menghasilkan pemeringkatan lima resume yang sesuai dengan hasil evaluasi pemeringkatan ahli terhadap deskripsi lowongan kerja ( $SRCC > 0,6$ ). Kemudian, pada skenario “Dengan Bobot”, Word2Vec dengan CosSim lebih unggul dalam menghasilkan pemeringkatan yang sesuai dengan evaluasi ahli. Walaupun begitu, Word2Vec dengan ISC pada skenario tanpa bobot *section*, memiliki nilai korelasi rata-rata 0,269565 yang dikategorikan sebagai korelasi lemah. Pada skenario dengan bobot *section*, Word2Vec dengan CosSim memiliki nilai korelasi rata-rata 0,183333 yang dikategorikan sebagai korelasi sangat lemah.

## 7.2 Saran

Berikut merupakan saran dari penelitian ini untuk penelitian berikutnya.

1. Pemberian bobot per *section* berdasarkan penilaian seorang ahli dapat menimbulkan bias, sehingga kurang mencerminkan variasi preferensi rekruter pada umumnya. Pada penelitian selanjutnya, disarankan untuk melibatkan lebih banyak ahli.
2. Meskipun implementasi *Improved Sqrt-Cosine* (ISC) *Similarity* dengan representasi teks Word2Vec unggul karena masih cukup mampu menangkap hubungan semantik kata kunci, tetapi kurang disarankan karena mengaburkan hubungan semantik asli. Jika berpacu pada seberapa relevan isian resume dengan kualifikasi lowongan kerja, maka implementasi ISC dengan TF-IDF lebih disarankan untuk digunakan.
3. Jumlah kualifikasi lowongan kerja dengan parameter-parameter yang di atas ambang batas rata-rata lebih sedikit dibandingkan yang di bawah ambang batas. Oleh karena itu, penelitian selanjutnya disarankan mencantumkan kata-kata teknis spesifik terkait suatu posisi pada deskripsi kualifikasi lowongan kerja agar deskripsi yang digunakan tidak terlalu umum.
4. Penelitian ini berfokus pada peringkat sebagai parameter utama, parameter relevansi dan senioritas ditambahkan sebagai parameter pendukung. Untuk penelitian selanjutnya, disarankan menganalisa hubungan parameter peringkat dengan parameter relevansi dan senioritas.
5. Pemrosesan kata untuk mengelompokkan tingkat senioritas pada setiap resume menjadi *entry level*, *junior level*, dan *management level* dapat meningkatkan kesesuaian resume-resume dari keluaran implementasi metode dengan tingkat senioritas yang dibutuhkan sesuai deskripsi kualifikasi lowongan kerja.
6. Pemrosesan ekstraksi *section* pada penelitian ini memanfaatkan kolom *Resume\_html* yang merupakan fitur asli dari *dataset* resume. Untuk penelitian selanjutnya, disarankan mengembangkan proses ekstraksi *section* tanpa memanfaatkan *class* dari struktur HTML.
7. Penyeragaman nama-nama *section* pada penelitian ini dilakukan secara manual. Untuk penelitian selanjutnya, disarankan mengembangkan proses penyeragaman nama-nama *section* secara semantik.
8. Penelitian ini hanya menggunakan *dataset* resume dan kualifikasi lowongan kerja dengan Bahasa Inggris. Untuk penelitian selanjutnya, disarankan mengembangkan *dataset* resume dan kualifikasi lowongan kerja dengan Bahasa Indonesia.

## DAFTAR REFERENSI

- Abdusyukur, F., 2023. *PENERAPAN ALGORITMA SUPPORT VECTOR MACHINE (SVM) UNTUK KLASIFIKASI PENCEMARAN NAMA BAIK DI MEDIA SOSIAL TWITTER*. KOMPUTA : Jurnal Ilmiah Komputer dan Informatika, 12(1), pp. 73-82.
- Alsharef, A., Sonia, Nassour, H. & Sharma, J., 2023. *Exploring the Efficiency of Text-Similarity Measures in Automated Resume Screening for Recruitment*. New Delhi, India, IEEE, pp. 36-42.
- Amin, M. D. et al., 2023. *Real Time Data based Automated Resume Classification and Job Matching using SVC, Jaccard Index and Cosine Similarity*. Roorkee, India, IEEE, pp. 1-6.
- Ayuningtyas, P. & Tantyoko, H., 2024. *Perbandingan Metode Word2vec Model Skipgram pada Ulasan Aplikasi Linkaja menggunakan Algoritma Bidirectional LSTM dan Support Vector Machine*. JUSTIN (Jurnal Sistem dan Teknologi Informasi), 12(1), pp. 189-196.
- Badan Pengembangan dan Pembinaan Bahasa, Kementerian Pendidikan, Kebudayaan, Riset, dan Teknologi Republik Indonesia, 2016. *KBBI VI Daring*. [Online]  
Tersedia di: <<https://kbbi.kemdikbud.go.id/entri/resume>>
- Bhawal, S., 2021. *Kaggle*. [Online]  
Tersedia di: <<https://www.kaggle.com/datasets/snehaanbhawal/resume-dataset>>
- Cambridge University Press & Assessment, 2024. *Meaning of curriculum vitae in English*. [Online]  
Tersedia di: <<https://dictionary.cambridge.org/dictionary/essential-british-english/curriculum-vitae>>
- Cholissodin, I. & Riyandani, E., 2018. *Big Data vs Big Information vs Big Knowledge*. Malang: Fakultas Ilmu Komputer Universitas Brawijaya.
- Cowley, H. P. et al., 2022. A framework for rigorous evaluation of human performance in human and machine learning comparison studies. *Scientific Reports*, 12(5444).
- Daryani, C. et al., 2020. AN AUTOMATED RESUME SCREENING SYSTEM USING NATURAL LANGUAGE PROCESSING AND SIMILARITY. *Ethics and Information Technology (ETIT)*, 2(2), pp. 99-103.
- Dewan Perwakilan Rakyat Republik Indonesia - Komisi IX, 2023. *Tingkat Pengangguran Terbuka Masih Jauh di Atas Target RPJMN*. [Online]  
Tersedia di: <<https://www.dpr.go.id/berita/detail/id/47507/t/Tingkat%20Pengangguran%20Terbuka%20Masih%20Jauh%20di%20Atas%20Target%20RPJMN>>

- Dwivedi, A. & Anand, S. K., 2023. *Word Embedding using Skip Gram Approach*. Interdisciplinary Journal of Contemporary Research, 10(3), pp. 1-5.
- Effendi, M. S., 2013. *Desain Eksperimental dalam Penelitian Pendidikan*. Jurnal Perspektif Pendidikan, 6(1), pp. 87-102.
- Guritno, S., S. & Rahardja, U., 2011. *Theory and Application of IT RESEARCH*. Penerbit Andi.
- He, Z., Dumdumaya, C. E. & Quimno, V. A., 2024. *MEASUREMENT OF SEMANTIC TEXT SIMILARITY*. Journal of Theoretical and Applied Information Technology, 102(5), pp. 1673-1685.
- H. & H., 2024. SPEARMAN'S RANK CORRELATION ANALYSIS METHOD TO IDENTIFY CHANGES IN THE GPA OF GRADUATES FROM THE 5TH BATCH OF THE TEACHING CAMPUS PROGRAM AT UNIVERSITAS BAKTI INDONESIA. *TRANSPIBLIKA INTERNATIONAL RESEARCH IN EXACT SCIENCES (Tires)*, 30 8, 3(3), pp. 18-27.
- International Monetary Fund, 2024. *World Economic Outlook (April 2024) - Unemployment Rate*. [Online] Tersedia di: <https://www.imf.org/external/datamapper/LUR@WEO/OEMDC/ADVEC/WEOWORLD/DA>
- Iskandar, D. & Kurniawan, A., 2025. ANALISIS PERBANDINGAN TEKNIK WORD2VEC DAN DOC2VECDALAM MENGIKUR KEMIRIPAN DOKUMEN MENGGUNAKAN COSINE SIMILARITY. *Jurnal Teknologi Informasi dan Ilmu Komputer (JTIK)*, 12(1), pp. 133-144.
- Jawale, D. S. et al., 2024. *COSINE SIMILARITY: A KEY DRIVER FOR ENHANCED RECOMMENDATION SYSTEMS*. International Research Journal of Modernization in Engineering Technology and Science, 06(04), pp. 1466-1470.
- Kementerian Ketenagakerjaan RI - Badan Perencanaan dan Pengembangan Ketenagakerjaan, 2021. *REVIEW RENCANA TENAGA KERJA NASIONAL 2020-2024*. [Online] Tersedia di: [https://satudata.kemnaker.go.id/satudata-public/2022/04/files/publikasi/1649938621648\\_Buku%2520Review%2520RTKN%202020%202024.pdf](https://satudata.kemnaker.go.id/satudata-public/2022/04/files/publikasi/1649938621648_Buku%2520Review%2520RTKN%202020%202024.pdf)
- Kulshretha, S. & Lodha, L., 2023. *Performance Evaluation of Word Embedding Algorithms*. International Journal of Innovative Science and Research Technology, 8(12), pp. 1555-1561.
- Kumaladewi, A. K., 2018. *EFEKTIVITAS REKRUTMEN DAN SELEKSI DALAM MEMENUHI KEBUTUHAN TENAGA PERAWAT DI RSIA MUSLIMAT JOMBA*. PARSIMONIA Jurnal Akuntansi, Manajemen, dan Bisnis, 11 4, 5(1), pp. 29-40.
- Lailasari, N. A. et al., 2024. *Pengaruh Pengangguran Terhadap Pertumbuhan Ekonomi*. IJM: Indonesian Journal of Multidisciplinary, 2(5), pp. 275-286.

- Musfiqon, 2016. *Panduan Lengkap Metodologi Penelitian Pendidikan*. PT. Prestasi Pustakaraya.
- Meyer, D., 2016. *How exactly does word2vec work?*
- P, A., K, A. K., Bharadwaj, S. K. & Venugopalan, M., 2024. *Semantic Similarity Analysis for Resume Filtering using PySpark*. Pune, India, IEEE, pp. 1-5
- Prasetya, D. D., Wibawa, A. P. & Hirashima, T., 2018. *The performance of text similarity algorithms*. International Journal of Advances in Intelligent Informatics, 4(1), pp. 63-69.
- Prasetya, M. A., Wulandari, M. & Nikmah, S. A., 2024. *Implementasi NLP(Natural Language Processing) Dasar pada Analisis Sentiment Review Spotify*. PROSIDING SEMINAR NASIONAL TEKNOLOGI DAN SAINS TAHUN 2024, pp. 145-153.
- Pundir, R. S. et al., 2024. *Enhancing Resume Recommendation System through Skill-based Similarity using Deep Learning Models*.
- Ramadhan, R. F., Wijoyo, S. H. & Saputra, M. C., 2023. *Penerapan Metode K-Means Clustering pada Ulasan Perumahan PT XYZ di Google Maps untuk Formulasi Strategi Bisnis dengan Analisis SWOT*. Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer, 7(6), pp. 2879-2888.
- Řehůřek, R., 2024. *Word2vec embeddings*. [Online] Tersedia di: <<https://radimrehurek.com/gensim/models/word2vec.html>>
- scikit-learn developers, 2025. *User Guide*. [Online] Tersedia di: <[https://scikit-learn.org/stable/modules/feature\\_extraction.html#text-feature-extraction](https://scikit-learn.org/stable/modules/feature_extraction.html#text-feature-extraction)>
- Septiani, D. & Isabela, I., 2022. *ANALISIS TERM FREQUENCY INVERSE DOCUMENT FREQUENCY (TF-IDF)*. SINTESIA: Jurnal Sistem dan Teknologi Informasi Indonesia, 01(2), pp. 81-88.
- Sihombing, D. O., 2022. *Implementasi Natural Language Processing (NLP) dan Algoritma Cosine Similarity dalam Penilaian Ujian Esai Otomatis*. Jurnal Sistem Komputer dan Informatika (JSON), 4(2), pp. 396-406.
- Sohangir, S. & Wang, D., 2017. *Improved sqrt-cosine similarity measurement*. Journal of Big Data, 4(25), pp. 1-13.
- Stanford Career Education, 2018. *Pursuing Meaningful Work: A Strategies Guide for PhDs and Postdocs*. [Online] Tersedia di: <[https://careered.stanford.edu/sites/g/files/sbiybj22801/files/media/file/stanfordphd\\_pmw\\_18-19.pdf](https://careered.stanford.edu/sites/g/files/sbiybj22801/files/media/file/stanfordphd_pmw_18-19.pdf)>
- Stanford Career Education, 2024. *Steps to Writing Your Resume*. [Online] Tersedia di: <<https://careered.stanford.edu/resources/resources-links#resume>>

- Suningsih, S. et al., 2024. *Pelatihan Pembuatan Curriculum Vitae dalam Bahasa Inggris yang Berbasis Application Tracking System*. Jurnal Nusantara Mengabdi, 3(2), pp. 85-93.
- Temizhan, E., Mirtagioglu, H. & Mendes, M., 2022. *Which Correlation Coefficient Should Be Used for Investigating Relations between Quantitative Variables?*. American Academic Scientific Research Journal for Engineering, Technology, and Sciences, 85(1), pp. 265-277.
- Titisari, M. & Ikhwan, K., 2021. *Proses Rekrutmen dan Seleksi: Potensi Ketidakefektifan dan Faktornya*. JMK (Jurnal Manajemen dan Kewirausahaan), 6(3), pp. 11-27.
- Wujarso, R., 2022. *PERAN HUMAN CAPITAL DALAM PERTUMBUHAN EKONOMI*. Journal of Information System, Applied, Management, Accounting and Research, 6(2), pp. 430-438.

## LAMPIRAN A SURAT PERNYATAAN VALIDITAS

### SURAT PERNYATAAN VALIDITAS

Yang bertanda tangan di bawah ini:

Nama : Rajiv Maulana

Jabatan : Talent Acquisition

Sebagai : Validator

Telah membaca instrumen penelitian dan memvalidasi kelayakan hasil evaluasi parameter-parameter pengujian yang digunakan dalam penelitian skripsi dengan judul "Implementasi Improved Sqrt-Cosine Similarity Untuk Pemeringkatan Resume Berdasarkan Kualifikasi Lowongan Kerja" oleh peneliti:

Nama : Khansa Salsabila Sangdiva Laksono

NIM : 215150201111068

Jurusan : Teknik Informatika

Program Studi : Teknik Informatika

Fakultas : Fakultas Ilmu Komputer Universitas Brawijaya

Demikian surat pernyataan ini dibuat untuk dipergunakan seperlunya.

Jakarta, 30 April 2025

Validator,



(Rajiv Maulana)

## LAMPIRAN B BOBOT PER SECTION BERDASARKAN INDUSTRI

### BOBOT PER SECTION BERDASARKAN INDUSTRI

Narasumber : Rajiv Maulana

Jabatan : *Talent Acquisition*

Industri	Section	Bobot
ENGINEERING	Summary	0
ENGINEERING	Accomplishments/Awards	10
ENGINEERING	Skills/Qualifications	15
ENGINEERING	Education	20
ENGINEERING	Experience	15
ENGINEERING	Organization	5
ENGINEERING	Projects	15
ENGINEERING	Certifications	15
ENGINEERING	Portfolio	10
HR	Summary	5
HR	Accomplishments/Awards	10
HR	Skills/Qualifications	15
HR	Education	20
HR	Experience	20
HR	Organization	10
HR	Projects	5
HR	Certifications	15
HR	Portfolio	5
AVIATION	Summary	0
AVIATION	Accomplishments/Awards	10
AVIATION	Skills/Qualifications	15
AVIATION	Education	30
AVIATION	Experience	15
AVIATION	Organization	0

Jakarta, 19 Februari 2025

  
(Rajiv Maulana)

<b>Industri</b>	<b>Section</b>	<b>Bobot</b>
AVIATION	Projects	0
AVIATION	Certifications	25
AVIATION	Portfolio	5
ACCOUNTANT	Summary	0
ACCOUNTANT	Accomplishments/Awards	5
ACCOUNTANT	Skills/Qualifications	25
ACCOUNTANT	Education	20
ACCOUNTANT	Experience	20
ACCOUNTANT	Organization	5
ACCOUNTANT	Projects	5
ACCOUNTANT	Certifications	15
ACCOUNTANT	Portfolio	5
DESIGNER	Summary	5
DESIGNER	Accomplishments/Awards	10
DESIGNER	Skills/Qualifications	20
DESIGNER	Education	15
DESIGNER	Experience	20
DESIGNER	Organization	5
DESIGNER	Projects	10
DESIGNER	Certifications	10
DESIGNER	Portfolio	20
ADVOCATE	Summary	5
ADVOCATE	Accomplishments/Awards	10
ADVOCATE	Skills/Qualifications	15
ADVOCATE	Education	20
ADVOCATE	Experience	15
ADVOCATE	Organization	5
ADVOCATE	Projects	10
ADVOCATE	Certifications	20

Jakarta, 19 Februari 2025



(Rajiv Maulana)

<b>Industri</b>	<b>Section</b>	<b>Bobot</b>
ADVOCATE	Portfolio	10
AGRICULTURE	Summary	0
AGRICULTURE	Accomplishments/Awards	10
AGRICULTURE	Skills/Qualifications	20
AGRICULTURE	Education	20
AGRICULTURE	Experience	30
AGRICULTURE	Organization	5
AGRICULTURE	Projects	10
AGRICULTURE	Certifications	10
AGRICULTURE	Portfolio	5
APPAREL	Summary	5
APPAREL	Accomplishments/Awards	10
APPAREL	Skills/Qualifications	10
APPAREL	Education	25
APPAREL	Experience	15
APPAREL	Organization	15
APPAREL	Projects	15
APPAREL	Certifications	15
APPAREL	Portfolio	15
ARTS	Summary	5
ARTS	Accomplishments/Awards	10
ARTS	Skills/Qualifications	10
ARTS	Education	10
ARTS	Experience	25
ARTS	Organization	10
ARTS	Projects	15
ARTS	Certifications	10
ARTS	Portfolio	15
AUTOMOBILE	Summary	0

Jakarta, 19 Februari 2025



(Rajiv Maulana)

<b>Industri</b>	<b>Section</b>	<b>Bobot</b>
AUTOMOBILE	Accomplishments/Awards	10
AUTOMOBILE	Skills/Qualifications	15
AUTOMOBILE	Education	20
AUTOMOBILE	Experience	20
AUTOMOBILE	Organization	10
AUTOMOBILE	Projects	10
AUTOMOBILE	Certifications	10
AUTOMOBILE	Portfolio	5
BANKING	Summary	0
BANKING	Accomplishments/Awards	15
BANKING	Skills/Qualifications	10
BANKING	Education	25
BANKING	Experience	25
BANKING	Organization	5
BANKING	Projects	10
BANKING	Certifications	10
BANKING	Portfolio	5
BPO	Summary	0
BPO	Accomplishments/Awards	15
BPO	Skills/Qualifications	10
BPO	Education	25
BPO	Experience	25
BPO	Organization	5
BPO	Projects	10
BPO	Certifications	10
BPO	Portfolio	5
BUSINESS-DEVELOPMENT	Summary	5
BUSINESS-DEVELOPMENT	Accomplishments/Awards	15
BUSINESS-DEVELOPMENT	Skills/Qualifications	20

Jakarta, 19 Februari 2025



(Rajiv Maulana)

<b>Industri</b>	<b>Section</b>	<b>Bobot</b>
BUSINESS-DEVELOPMENT	Education	15
BUSINESS-DEVELOPMENT	Experience	20
BUSINESS-DEVELOPMENT	Organization	5
BUSINESS-DEVELOPMENT	Projects	10
BUSINESS-DEVELOPMENT	Certifications	10
BUSINESS-DEVELOPMENT	Portfolio	5
<b>CHEF</b>	Summary	5
CHEF	Accomplishments/Awards	15
CHEF	Skills/Qualifications	20
CHEF	Education	20
CHEF	Experience	25
CHEF	Organization	5
CHEF	Projects	5
CHEF	Certifications	15
CHEF	Portfolio	10
<b>CONSTRUCTION</b>	Summary	0
CONSTRUCTION	Accomplishments/Awards	10
CONSTRUCTION	Skills/Qualifications	15
CONSTRUCTION	Education	20
CONSTRUCTION	Experience	15
CONSTRUCTION	Organization	5
CONSTRUCTION	Projects	15
CONSTRUCTION	Certifications	15
CONSTRUCTION	Portfolio	10
<b>CONSULTANT</b>	Summary	5
CONSULTANT	Accomplishments/Awards	15
CONSULTANT	Skills/Qualifications	20
CONSULTANT	Education	15
CONSULTANT	Experience	20

Jakarta, 19 Februari 2025



(Rajiv Maulana)

<b>Industri</b>	<b>Section</b>	<b>Bobot</b>
CONSULTANT	Organization	5
CONSULTANT	Projects	10
CONSULTANT	Certifications	10
CONSULTANT	Portfolio	5
DIGITAL-MEDIA	Summary	5
DIGITAL-MEDIA	Accomplishments/Awards	10
DIGITAL-MEDIA	Skills/Qualifications	10
DIGITAL-MEDIA	Education	10
DIGITAL-MEDIA	Experience	25
DIGITAL-MEDIA	Organization	10
DIGITAL-MEDIA	Projects	15
DIGITAL-MEDIA	Certifications	10
DIGITAL-MEDIA	Portfolio	15
FINANCE	Summary	0
FINANCE	Accomplishments/Awards	15
FINANCE	Skills/Qualifications	10
FINANCE	Education	25
FINANCE	Experience	25
FINANCE	Organization	5
FINANCE	Projects	10
FINANCE	Certifications	10
FINANCE	Portfolio	5
FITNESS	Summary	5
FITNESS	Accomplishments/Awards	10
FITNESS	Skills/Qualifications	15
FITNESS	Education	10
FITNESS	Experience	30
FITNESS	Organization	10
FITNESS	Projects	5

Jakarta, 19 Februari 2025



(Rajiv Maulana)

<b>Industri</b>	<b>Section</b>	<b>Bobot</b>
FITNESS	Certifications	15
FITNESS	Portfolio	5
HEALTHCARE	Summary	0
HEALTHCARE	Accomplishments/Awards	10
HEALTHCARE	Skills/Qualifications	15
HEALTHCARE	Education	30
HEALTHCARE	Experience	15
HEALTHCARE	Organization	5
HEALTHCARE	Projects	5
HEALTHCARE	Certifications	15
HEALTHCARE	Portfolio	5
INFORMATION-TECHNOLOGY	Summary	0
INFORMATION-TECHNOLOGY	Accomplishments/Awards	10
INFORMATION-TECHNOLOGY	Skills/Qualifications	15
INFORMATION-TECHNOLOGY	Education	20
INFORMATION-TECHNOLOGY	Experience	15
INFORMATION-TECHNOLOGY	Organization	5
INFORMATION-TECHNOLOGY	Projects	15
INFORMATION-TECHNOLOGY	Certifications	15
INFORMATION-TECHNOLOGY	Portfolio	10
PUBLIC-RELATIONS	Summary	5
PUBLIC-RELATIONS	Accomplishments/Awards	15
PUBLIC-RELATIONS	Skills/Qualifications	20
PUBLIC-RELATIONS	Education	15

Jakarta, 19 Februari 2025



(Rajiv Maulana)

<b>Industri</b>	<b>Section</b>	<b>Bobot</b>
PUBLIC-RELATIONS	Experience	20
PUBLIC-RELATIONS	Organization	5
PUBLIC-RELATIONS	Projects	10
PUBLIC-RELATIONS	Certifications	10
PUBLIC-RELATIONS	Portfolio	5
SALES	Summary	0
SALES	Accomplishments/Awards	15
SALES	Skills/Qualifications	25
SALES	Education	15
SALES	Experience	25
SALES	Organization	5
SALES	Projects	5
SALES	Certifications	5
SALES	Portfolio	5
TEACHER	Summary	5
TEACHER	Accomplishments/Awards	15
TEACHER	Skills/Qualifications	20
TEACHER	Education	20
TEACHER	Experience	20
TEACHER	Organization	5
TEACHER	Projects	5
TEACHER	Certifications	10
TEACHER	Portfolio	10

Jakarta, 19 Februari 2025



(Rajiv Maulana)

## **LAMPIRAN C HASIL PEMERINGKATAN LIMA RESUME PER KUALIFIKASI LOWONGAN KERJA**

### **C.1 Tanpa Bobot - TF-IDF dan *Improved Sqrt-Cosine Similarity***

<b>Rank</b>	<b>Resume ID</b>	<b>Position</b>	<b>Similarity Score</b>	<b>Rank Expert</b>	<b>Relevance</b>	<b>Seniority</b>
1	38688388	Business Development Executive	0,1919605 416	3	TRUE	FALSE
2	31638814	Business Development Executive	0,1863253 744	2	TRUE	FALSE
3	18311419	Business Development Executive	0,1783844 987	5	TRUE	FALSE
4	15535920	Business Development Executive	0,1708420 879	1	TRUE	FALSE
5	17132168	Business Development Executive	0,1682773 892	4	TRUE	FALSE
1	26932091	CLUB GENERAL MANAGER	0,1871587 776	5	TRUE	TRUE
2	17818707	CLUB GENERAL MANAGER	0,1794173 858	2	TRUE	TRUE
3	15535920	CLUB GENERAL MANAGER	0,1714170 333	4	TRUE	TRUE
4	38688388	CLUB GENERAL MANAGER	0,1605149 45	3	TRUE	TRUE

5	25162378	CLUB GENERAL MANAGE R	0,1561663 095	1	TRUE	TRUE
1	27246366	Construct ion Superviso r	0,2309432 449	2	TRUE	TRUE
2	39027764	Construct ion Superviso r	0,2258310 516	1	TRUE	TRUE
3	12839152	Construct ion Superviso r	0,2158709 047	3	TRUE	TRUE
4	22718826	Construct ion Superviso r	0,2089757 732	4	TRUE	FALSE
5	26994282	Construct ion Superviso r	0,1991065 198	5	FALSE	TRUE
1	68781345	Creative Director / Manager	0,1449183 029	3	TRUE	TRUE
2	13964744	Creative Director / Manager	0,1365746 594	5	FALSE	FALSE
3	308648 28	Creative Director / Manager	0,1303657 835	1	TRUE	TRUE
4	17781039	Creative Director / Manager	0,1282461 853	4	FALSE	FALSE
5	22706174	Creative Director / Manager	0,1264199 202	2	TRUE	TRUE
1	22754014	Digital and Social Media Executive	0,1351744 084	4	FALSE	TRUE

2	16620172	Digital and Social Media Executive	0,1303300 184	1	TRUE	TRUE
3	18905648	Digital and Social Media Executive	0,1297599	5	FALSE	TRUE
4	18927233	Digital and Social Media Executive	0,1290032 048	3	TRUE	FALSE
5	16536141	Digital and Social Media Executive	0,1178854 437	2	TRUE	FALSE
1	14937492	Digital Banking Officer	0,1692186 127	4	FALSE	FALSE
2	27080812	Digital Banking Officer	0,1653474 431	3	FALSE	FALSE
3	26932091	Digital Banking Officer	0,1628028 08	5	FALSE	FALSE
4	22423839	Digital Banking Officer	0,1595269 82	1	FALSE	FALSE
5	25038571	Digital Banking Officer	0,1542614 74	2	FALSE	FALSE
1	34252537	Executive Chef	0,2393743 131	1	TRUE	TRUE
2	29775391	Executive Chef	0,2362169 724	4	TRUE	TRUE
3	10653119	Executive Chef	0,2324195 106	2	TRUE	TRUE
4	25924968	Executive Chef	0,2310007 976	5	TRUE	TRUE
5	16924102	Executive Chef	0,2229271 471	3	TRUE	TRUE

1	21338490	Finance Executive / Accountant	0,2365662907	5	TRUE	FALSE
2	25846894	Finance Executive / Accountant	0,2339642328	4	TRUE	FALSE
3	25862026	Finance Executive / Accountant	0,1916905061	3	TRUE	TRUE
4	29999135	Finance Executive / Accountant	0,1887091447	1	TRUE	TRUE
5	28969385	Finance Executive / Accountant	0,1883069276	2	TRUE	TRUE
1	23734441	Finance Officer (Jr/Sr.)	0,2471592921	3	TRUE	TRUE
2	28298773	Finance Officer (Jr/Sr.)	0,2255675767	1	TRUE	TRUE
3	29999135	Finance Officer (Jr/Sr.)	0,2112704543	2	TRUE	TRUE
4	53640713	Finance Officer (Jr/Sr.)	0,1999300064	4	TRUE	FALSE
5	21338490	Finance Officer (Jr/Sr.)	0,1984463036	5	TRUE	FALSE
1	38946032	Financial Consolidation	0,1817793037	4	TRUE	TRUE

		Consultant				
2	70541112	Financial Consolidation Consultant	0,1811155551	1	TRUE	TRUE
3	21338490	Financial Consolidation Consultant	0,1745877097	5	TRUE	TRUE
4	29821051	Financial Consolidation Consultant	0,1730965599	2	TRUE	TRUE
5	25862026	Financial Consolidation Consultant	0,1725639591	3	TRUE	TRUE
1	18354623	Graphics Designer	0,2102328956	4	TRUE	FALSE
2	18460045	Graphics Designer	0,1993908853	2	TRUE	TRUE
3	20210676	Graphics Designer	0,1793144558	5	FALSE	FALSE
4	22560013	Graphics Designer	0,1749108248	1	TRUE	TRUE
5	26046064	Graphics Designer	0,1747824455	3	TRUE	TRUE
1	30862904	HR Specialist	0,2428395272	1	TRUE	TRUE
2	24508725	HR Specialist	0,2396608996	2	TRUE	TRUE
3	16877897	HR Specialist	0,2301432719	5	FALSE	TRUE
4	11480899	HR Specialist	0,2216053793	3	FALSE	TRUE
5	53701275	HR Specialist	0,2189996296	4	FALSE	TRUE

1	39413067	INFORMATION & TECHNOLOGY STAFF	0,2356860 514	1	TRUE	TRUE
2	17983957	INFORMATION & TECHNOLOGY STAFF	0,2273222 616	4	FALSE	FALSE
3	91635250	INFORMATION & TECHNOLOGY STAFF	0,2184461 588	3	TRUE	FALSE
4	15535920	INFORMATION & TECHNOLOGY STAFF	0,2094339 334	5	FALSE	FALSE
5	36434348	INFORMATION & TECHNOLOGY STAFF	0,1971253 698	2	TRUE	FALSE
1	36671891	Junior Associate Lawyer	0,1206467 025	3	FALSE	FALSE
2	19557384	Junior Associate Lawyer	0,1203686 863	4	FALSE	FALSE
3	10332998	Junior Associate Lawyer	0,1195592 716	1	TRUE	TRUE
4	15100547	Junior Associate Lawyer	0,1180229 251	2	TRUE	TRUE
5	11065180	Junior Associate Lawyer	0,1170764 78	5	FALSE	FALSE
1	23719943	Junior Designer for Apparel	0,1771679 315	2	TRUE	FALSE

2	15746146	Junior Designer for Apparel	0,1627493 142	1	TRUE	TRUE
3	26503829	Junior Designer for Apparel	0,1479395 386	3	TRUE	FALSE
4	12122372	Junior Designer for Apparel	0,1410829 334	4	FALSE	FALSE
5	26932091	Junior Designer for Apparel	0,1367025 452	5	FALSE	FALSE
1	26932091	Manager Aviation Safety, Quality and Security	0,2015879 825	5	FALSE	FALSE
2	11169163	Manager Aviation Safety, Quality and Security	0,2013763 559	2	TRUE	TRUE
3	19796840	Manager Aviation Safety, Quality and Security	0,2010987 541	4	FALSE	FALSE
4	28186635	Manager Aviation Safety, Quality and Security	0,1877334 005	3	TRUE	FALSE
5	28383893	Manager Aviation Safety, Quality	0,1862704 602	1	TRUE	TRUE

		and Security				
1	16356151	Medical Doctor	0,2157386 58	1	FALSE	TRUE
2	13565152	Medical Doctor	0,1875315 396	5	FALSE	FALSE
3	17818707	Medical Doctor	0,1628595 03	4	FALSE	FALSE
4	12544735	Medical Doctor	0,1595774 142	2	FALSE	FALSE
5	43994605	Medical Doctor	0,1595204 988	3	FALSE	FALSE
1	77828437	Productio n Engineeri ng	0,1264697 128	1	TRUE	TRUE
2	55595908	Productio n Engineeri ng	0,1201321 813	5	FALSE	FALSE
3	28803888	Productio n Engineeri ng	0,1154167 25	3	TRUE	FALSE
4	30288581	Productio n Engineeri ng	0,1118393 979	4	FALSE	FALSE
5	86828820	Productio n Engineeri ng	0,1101805 298	2	FALSE	FALSE
1	21297828	Public Relations Officer	0,1907608 096	3	TRUE	TRUE
2	13129275	Public Relations Officer	0,1802253 973	5	FALSE	TRUE
3	28290448	Public Relations Officer	0,1736107 376	2	TRUE	TRUE

4	31220062	Public Relations Officer	0,1732900 471	4	FALSE	TRUE
5	20210676	Public Relations Officer	0,1725535 861	1	TRUE	TRUE
1	26888302	Quality Control Supervisor - Corn Commodity	0,1252811 704	2	FALSE	TRUE
2	26932091	Quality Control Supervisor - Corn Commodity	0,1239398 512	4	FALSE	TRUE
3	20905088	Quality Control Supervisor - Corn Commodity	0,1106055 359	3	FALSE	FALSE
4	28020046	Quality Control Supervisor - Corn Commodity	0,1081390 286	1	FALSE	TRUE
5	22861181	Quality Control Supervisor - Corn Commodity	0,1069757 33	5	FALSE	FALSE
1	26932091	Regional Sales Manager	0,2134135 079	5	FALSE	TRUE
2	25038571	Regional Sales Manager	0,2032024 859	2	TRUE	TRUE
3	27080812	Regional Sales Manager	0,2029929 331	3	FALSE	TRUE

4	38688388	Regional Sales Manager	0,1964119 402	4	FALSE	FALSE
5	17818707	Regional Sales Manager	0,1949221 894	1	TRUE	TRUE
1	35474904	Spare part Admin	0,1420329 913	5	FALSE	FALSE
2	22861181	Spare part Admin	0,1379583 676	4	FALSE	FALSE
3	19473948	Spare part Admin	0,1371374 364	1	FALSE	TRUE
4	71772815	Spare part Admin	0,1371227 976	2	FALSE	TRUE
5	24670867	Spare part Admin	0,1369340 491	3	FALSE	FALSE
1	15850434	Teachers	0,2201057 854	2	TRUE	TRUE
2	96547039	Teachers	0,1978795 776	1	TRUE	TRUE
3	28772892	Teachers	0,1889872 465	3	TRUE	TRUE
4	58105060	Teachers	0,1869135 69	5	TRUE	TRUE
5	37220856	Teachers	0,1804859 322	4	FALSE	TRUE
1	26932091	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,2280308 996	1	TRUE	TRUE
2	68781345	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,2043528 088	5	FALSE	FALSE

3	17818707	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,1783846 929	4	FALSE	FALSE
4	27884470	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,1726401 754	3	FALSE	FALSE
5	38688388	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,1670092 401	2	TRUE	FALSE

## C.2 Tanpa Bobot - Word2Vec dan Cosine Similarity

Rank	Resume ID	Position	Similarity Score	Rank Expert	Relevance	Seniority
1	26932091	Business Development Executive	0,8910836 577	4	TRUE	FALSE
2	10464113	Business Development Executive	0,8697264 314	3	TRUE	FALSE
3	13352113	Business Development Executive	0,8397827 595	5	TRUE	TRUE
4	17132168	Business Development Executive	0,8354100 585	1	TRUE	TRUE

5	27715131	Business Development Executive	0,8342830539	2	TRUE	FALSE
1	26932091	CLUB GENERAL MANAGER	0,8795587818	5	FALSE	FALSE
2	10464113	CLUB GENERAL MANAGER	0,8576420188	4	FALSE	FALSE
3	13411858	CLUB GENERAL MANAGER	0,8569679499	3	FALSE	FALSE
4	27715131	CLUB GENERAL MANAGER	0,8542478204	2	FALSE	TRUE
5	12938389	CLUB GENERAL MANAGER	0,8491947293	1	TRUE	TRUE
1	27246366	Construction Supervisor	0,8558241725	1	TRUE	TRUE
2	39027764	Construction Supervisor	0,8522567153	4	TRUE	FALSE
3	26932091	Construction Supervisor	0,8479026556	5	FALSE	FALSE
4	16203589	Construction Supervisor	0,8425265431	3	FALSE	TRUE
5	12839152	Construction Supervisor	0,8412288904	2	TRUE	TRUE

1	13115648	Creative Director / Manager	0,8481530 845	3	FALSE	FALSE
2	81508860	Creative Director / Manager	0,8427884 43	1	TRUE	TRUE
3	129383 89	Creative Director / Manager	0,8389229 774	4	FALSE	FALSE
4	23917826	Creative Director / Manager	0,8372398 466	2	FALSE	FALSE
5	24588864	Creative Director / Manager	0,8319416 642	5	FALSE	FALSE
1	18905648	Digital and Social Media Executive	0,8833178 282	4	FALSE	TRUE
2	18354623	Digital and Social Media Executive	0,8470618 427	2	TRUE	FALSE
3	70750649	Digital and Social Media Executive	0,8403711 468	1	TRUE	FALSE
4	26932091	Digital and Social Media Executive	0,8330877 423	5	FALSE	FALSE
5	22754014	Digital and Social Media Executive	0,8241794 109	3	FALSE	TRUE
1	26932091	Digital Banking Officer	0,8829766 711	4	FALSE	TRUE
2	98379112	Digital Banking Officer	0,8419138 193	1	FALSE	FALSE

3	26167298	Digital Banking Officer	0,8403331 637	3	FALSE	FALSE
4	18824120	Digital Banking Officer	0,8402394 056	2	FALSE	TRUE
5	10464113	Digital Banking Officer	0,8382170 916	5	FALSE	TRUE
1	35579812	Executive Chef	0,8825965 822	5	TRUE	TRUE
2	29775391	Executive Chef	0,8810187 817	3	TRUE	TRUE
3	34252537	Executive Chef	0,8682833 076	4	TRUE	TRUE
4	21060367	Executive Chef	0,8642790 198	1	TRUE	TRUE
5	25924968	Executive Chef	0,8607410 938	2	TRUE	TRUE
1	20393721	Finance Executive / Accountant	0,8786240 816	3	TRUE	TRUE
2	23636277	Finance Executive / Accountant	0,8532985 747	2	TRUE	TRUE
3	70541112	Finance Executive / Accountant	0,8524324 417	1	TRUE	TRUE
4	22861181	Finance Executive / Accountant	0,8475373 387	5	FALSE	FALSE
5	26695839	Finance Executive /	0,8448839 585	4	FALSE	TRUE

		Accountant				
1	20393721	Finance Officer (Jr/Sr.)	0,8747340639	3	TRUE	FALSE
2	23734441	Finance Officer (Jr/Sr.)	0,8486161629	1	TRUE	TRUE
3	27558837	Finance Officer (Jr/Sr.)	0,8390957117	2	TRUE	TRUE
4	25497147	Finance Officer (Jr/Sr.)	0,8381744027	4	FALSE	TRUE
5	53640713	Finance Officer (Jr/Sr.)	0,8285536528	5	FALSE	TRUE
1	70541112	Financial Consolidation Consultant	0,8601319313	1	TRUE	TRUE
2	18365443	Financial Consolidation Consultant	0,8497776538	2	FALSE	TRUE
3	269320 91	Financial Consolidation Consultant	0,8467198412	4	FALSE	TRUE
4	26695839	Financial Consolidation Consultant	0,8388730884	5	FALSE	FALSE
5	19446337	Financial Consolidation Consultant	0,8272504449	3	FALSE	TRUE
1	18354623	Graphics Designer	0,8890629709	3	TRUE	FALSE

2	22754014	Graphics Designer	0,8265473 545	2	TRUE	TRUE
3	70750649	Graphics Designer	0,8213165 402	4	FALSE	FALSE
4	37664296	Graphics Designer	0,8205203 891	1	TRUE	TRUE
5	22848179	Graphics Designer	0,8096640 706	5	FALSE	FALSE
1	16877897	HR Specialist	0,9091145 694	2	FALSE	TRUE
2	26932091	HR Specialist	0,8979627 291	5	FALSE	FALSE
3	30862904	HR Specialist	0,8975547 701	1	TRUE	TRUE
4	29134372	HR Specialist	0,8676508 904	3	FALSE	FALSE
5	11289482	HR Specialist	0,8629413 128	4	FALSE	FALSE
1	26932091	INFORMATION & TECHNOLOGY STAFF	0,8692647 02	5	FALSE	TRUE
2	28672970	INFORMATION & TECHNOLOGY STAFF	0,8644542 694	3	TRUE	FALSE
3	10840430	INFORMATION & TECHNOLOGY STAFF	0,8521312 773	2	TRUE	FALSE
4	11957080	INFORMATION & TECHNOLOGY STAFF	0,8495982 438	1	TRUE	TRUE
5	15535920	INFORMATION & TECHNOLOGY STAFF	0,8477472 961	4	FALSE	FALSE

1	81508860	Junior Associate Lawyer	0,8411479 443	5	FALSE	FALSE
2	98379112	Junior Associate Lawyer	0,8350052 088	4	FALSE	FALSE
3	18297650	Junior Associate Lawyer	0,8287451 416	1	FALSE	FALSE
4	22485475	Junior Associate Lawyer	0,8282045 275	2	FALSE	FALSE
5	16877897	Junior Associate Lawyer	0,8276658 654	3	FALSE	FALSE
1	26932091	Junior Designer for Apparel	0,8629954 656	5	FALSE	FALSE
2	15154822	Junior Designer for Apparel	0,8216761 589	1	TRUE	FALSE
3	19195747	Junior Designer for Apparel	0,8122268 17	3	FALSE	FALSE
4	23719943	Junior Designer for Apparel	0,8101501 614	2	FALSE	FALSE
5	18354623	Junior Designer for Apparel	0,8057058 156	4	FALSE	FALSE
1	13195436	Manager Aviation Safety, Quality and Security	0,8760108 203	1	TRUE	TRUE
2	11169163	Manager Aviation Safety,	0,8636670 262	2	TRUE	TRUE

		Quality and Security				
3	12654876	Manager Aviation Safety, Quality and Security	0,8354426 771	4	FALSE	TRUE
4	35651876	Manager Aviation Safety, Quality and Security	0,8326146 305	3	FALSE	FALSE
5	26932091	Manager Aviation Safety, Quality and Security	0,8321723 739	5	FALSE	FALSE
1	25328428	Medical Doctor	0,8549853 325	2	FALSE	FALSE
2	96260484	Medical Doctor	0,8511528 522	1	FALSE	TRUE
3	37402097	Medical Doctor	0,8448984 325	5	FALSE	FALSE
4	15958967	Medical Doctor	0,8384091 616	3	FALSE	FALSE
5	12544735	Medical Doctor	0,8357500 881	4	FALSE	FALSE
1	54100393	Production Engineering	0,7940523 922	5	FALSE	FALSE
2	30288581	Production Engineering	0,7927010 179	2	FALSE	FALSE
3	12011623	Production Engineering	0,7900994 569	1	TRUE	TRUE

4	37751611	Production Engineering	0,7888528109	4	FALSE	FALSE
5	10751444	Production Engineering	0,7884584188	3	FALSE	FALSE
1	28290448	Public Relations Officer	0,8557876945	1	TRUE	TRUE
2	70750649	Public Relations Officer	0,8394137621	2	TRUE	FALSE
3	22861181	Public Relations Officer	0,8321976066	5	FALSE	FALSE
4	22754014	Public Relations Officer	0,8313117474	3	TRUE	TRUE
5	18354623	Public Relations Officer	0,8215153068	4	TRUE	FALSE
1	26932091	Quality Control Supervisor - Corn Commodity	0,8515201807	5	FALSE	TRUE
2	35651876	Quality Control Supervisor - Corn Commodity	0,8261224329	2	FALSE	FALSE
3	12011623	Quality Control Supervisor - Corn Commodity	0,8161604106	3	FALSE	FALSE
4	26070334	Quality Control Supervisor - Corn	0,8140973449	1	FALSE	TRUE

		Commodity				
5	26888302	Quality Control Supervisor - Corn Commodity	0,8107827604	4	FALSE	FALSE
1	26932091	Regional Sales Manager	0,913558503	3	FALSE	TRUE
2	28867567	Regional Sales Manager	0,8628078401	1	TRUE	TRUE
3	12059198	Regional Sales Manager	0,8436546087	2	FALSE	TRUE
4	18368613	Regional Sales Manager	0,8355645984	4	TRUE	FALSE
5	23917826	Regional Sales Manager	0,8340500742	5	FALSE	TRUE
1	26932091	Spare part Admin	0,8606192668	5	FALSE	TRUE
2	22861181	Spare part Admin	0,85548307	3	FALSE	FALSE
3	16378091	Spare part Admin	0,8486716747	1	FALSE	TRUE
4	23917826	Spare part Admin	0,8374442756	4	FALSE	TRUE
5	37764298	Spare part Admin	0,8364215493	2	FALSE	TRUE
1	28772892	Teachers	0,9286252558	4	TRUE	TRUE
2	15850434	Teachers	0,9054198861	2	TRUE	TRUE
3	54100393	Teachers	0,9002476335	5	TRUE	TRUE
4	37220856	Teachers	0,895565021	3	FALSE	TRUE

5	96547039	Teachers	0,8829045 147	1	TRUE	TRUE
1	26932091	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,9265301 426	1	TRUE	TRUE
2	11289482	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,8595816 135	2	TRUE	FALSE
3	16877897	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,8374321 014	5	FALSE	FALSE
4	30862904	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,8252724 558	4	FALSE	FALSE
5	12938389	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,8237882 972	3	FALSE	FALSE

### C.3 Tanpa Bobot - Word2Vec dan *Improved Sqrt-Cosine Similarity*

<i>Rank</i>	<i>Resume ID</i>	<i>Position</i>	<i>Similarity Score</i>	<i>Rank Expert</i>	<i>Relevance</i>	<i>Seniority</i>
1	26932091	Business Development Executive	0,9526863 545	2	TRUE	FALSE
2	10464113	Business Development Executive	0,9469295 215	5	FALSE	TRUE
3	27715131	Business Development Executive	0,9431682 632	4	TRUE	FALSE
4	91467795	Business Development Executive	0,9431253 452	1	TRUE	TRUE
5	17132168	Business Development Executive	0,9415056 337	3	TRUE	TRUE
1	26932091	CLUB GENERAL MANAGER	0,9499713 215	1	TRUE	TRUE
2	10464113	CLUB GENERAL MANAGER	0,9470196 187	5	FALSE	FALSE
3	13411858	CLUB GENERAL MANAGER	0,9465261 069	4	FALSE	TRUE
4	27715131	CLUB GENERAL MANAGER	0,9463747 964	3	FALSE	TRUE
5	12938389	CLUB GENERAL MANAGER	0,9449584 126	2	FALSE	TRUE

1	39027764	Construction Supervisor	0,9464103 571	4	TRUE	FALSE
2	27246366	Construction Supervisor	0,9418411 329	2	TRUE	TRUE
3	12839152	Construction Supervisor	0,9405705 386	1	TRUE	TRUE
4	26932091	Construction Supervisor	0,9388270 484	5	FALSE	TRUE
5	10176013	Construction Supervisor	0,9355313 332	3	TRUE	TRUE
1	22861181	Creative Director / Manager	0,9488284 06	2	FALSE	TRUE
2	13115648	Creative Director / Manager	0,9476442 985	5	FALSE	FALSE
3	815088 60	Creative Director / Manager	0,9465295 591	1	TRUE	TRUE
4	16899268	Creative Director / Manager	0,9464348 445	3	FALSE	TRUE
5	26932091	Creative Director / Manager	0,9457878 553	4	FALSE	TRUE
1	18905648	Digital and Social Media Executive	0,9541787 533	1	TRUE	TRUE
2	70750649	Digital and Social Media Executive	0,9476443 214	3	TRUE	FALSE

3	18354623	Digital and Social Media Executive	0,9467136 247	2	TRUE	FALSE
4	34712719	Digital and Social Media Executive	0,9436558 5	5	FALSE	TRUE
5	14304010	Digital and Social Media Executive	0,9412760 663	4	TRUE	TRUE
1	26932091	Digital Banking Officer	0,9548594 827	4	FALSE	FALSE
2	18905648	Digital Banking Officer	0,9489372 807	2	FALSE	FALSE
3	13352113	Digital Banking Officer	0,9462483 878	3	FALSE	TRUE
4	10464113	Digital Banking Officer	0,9460302 848	5	FALSE	TRUE
5	15423153	Digital Banking Officer	0,9442921 952	1	FALSE	TRUE
1	35579812	Executive Chef	0,9570838 092	5	FALSE	TRUE
2	29775391	Executive Chef	0,9524580 434	2	TRUE	TRUE
3	10276858	Executive Chef	0,9479883 343	4	TRUE	FALSE
4	16924102	Executive Chef	0,9479701 332	3	TRUE	TRUE
5	65373280	Executive Chef	0,9462183 423	1	TRUE	TRUE
1	20393721	Finance Executive / Accountant	0,9511945 992	3	FALSE	TRUE

2	70541112	Finance Executive / Accountant	0,9456473 07	1	TRUE	TRUE
3	25846894	Finance Executive / Accountant	0,9446961 732	2	TRUE	TRUE
4	19446337	Finance Executive / Accountant	0,9443189 041	4	FALSE	TRUE
5	25497147	Finance Executive / Accountant	0,9438716 017	5	FALSE	TRUE
1	20393721	Finance Officer ( Jr/Sr.)	0,9600386 8	3	FALSE	TRUE
2	23734441	Finance Officer ( Jr/Sr.)	0,9522399 562	2	TRUE	TRUE
3	25497147	Finance Officer ( Jr/Sr.)	0,9517891 475	4	FALSE	TRUE
4	53640713	Finance Officer ( Jr/Sr.)	0,9508016 156	5	FALSE	TRUE
5	34816637	Finance Officer ( Jr/Sr.)	0,9463608 681	1	TRUE	TRUE
1	70541112	Financial Consolidation Consultant	0,9417991 803	2	TRUE	TRUE
2	27330027	Financial Consolidation	0,9386546 765	4	FALSE	FALSE

		Consultant				
3	153632 77	Financial Consolidation Consultant	0,9359328 764	3	TRUE	TRUE
4	19446337	Financial Consolidation Consultant	0,9342442 62	1	TRUE	TRUE
5	18365443	Financial Consolidation Consultant	0,9331100 827	5	FALSE	TRUE
1	18354623	Graphics Designer	0,9568201 853	1	TRUE	FALSE
2	70750649	Graphics Designer	0,9531553 81	4	TRUE	FALSE
3	22754014	Graphics Designer	0,9445030 836	2	TRUE	FALSE
4	28679359	Graphics Designer	0,9435436 506	5	TRUE	FALSE
5	14304010	Graphics Designer	0,9426000 143	3	TRUE	FALSE
1	16877897	HR Specialist	0,9579303 028	1	TRUE	FALSE
2	30862904	HR Specialist	0,9561192 703	2	TRUE	FALSE
3	26932091	HR Specialist	0,9515269 99	5	FALSE	FALSE
4	22861181	HR Specialist	0,9453664 87	3	FALSE	FALSE
5	11289482	HR Specialist	0,9444434 517	4	FALSE	FALSE
1	10840430	INFORMATION & TECHNOLOGY STAFF	0,9484859 45	2	TRUE	FALSE

2	11676151	INFORMATION & TECHNOLOGY STAFF	0,9474958 077	1	TRUE	TRUE
3	26932091	INFORMATION & TECHNOLOGY STAFF	0,9460883 964	5	FALSE	FALSE
4	28672970	INFORMATION & TECHNOLOGY STAFF	0,9457937 874	3	TRUE	FALSE
5	17963031	INFORMATION & TECHNOLOGY STAFF	0,9457588 681	4	FALSE	FALSE
1	64589506	Junior Associate Lawyer	0,9402469 427	1	FALSE	FALSE
2	26932091	Junior Associate Lawyer	0,9400253 197	2	FALSE	FALSE
3	22861181	Junior Associate Lawyer	0,9393440 608	3	FALSE	FALSE
4	16877897	Junior Associate Lawyer	0,9382883 659	4	FALSE	FALSE
5	19557384	Junior Associate Lawyer	0,9380632 209	5	FALSE	FALSE
1	23719943	Junior Designer for Apparel	0,9404868 792	2	FALSE	FALSE
2	26932091	Junior Designer for Apparel	0,9394543 595	5	FALSE	FALSE

3	23917826	Junior Designer for Apparel	0,9328362 478	4	FALSE	FALSE
4	15154822	Junior Designer for Apparel	0,9310774 192	1	TRUE	FALSE
5	70750649	Junior Designer for Apparel	0,9298330 345	3	FALSE	FALSE
1	13195436	Manager Aviation Safety, Quality and Security	0,9486718 093	1	TRUE	TRUE
2	11169163	Manager Aviation Safety, Quality and Security	0,9479157 24	3	TRUE	TRUE
3	12654876	Manager Aviation Safety, Quality and Security	0,9438178 526	4	FALSE	FALSE
4	35651876	Manager Aviation Safety, Quality and Security	0,9438160 536	5	FALSE	FALSE
5	17483843	Manager Aviation Safety, Quality and Security	0,9411606 041	2	TRUE	TRUE
1	96260484	Medical Doctor	0,9365936 609	1	FALSE	TRUE

2	37402097	Medical Doctor	0,9359483 901	3	FALSE	TRUE
3	25328428	Medical Doctor	0,9357925 524	5	FALSE	TRUE
4	15499825	Medical Doctor	0,9343298 543	4	FALSE	TRUE
5	14667957	Medical Doctor	0,9322447 061	2	FALSE	TRUE
1	13087952	Production Engineering	0,9391430 121	2	FALSE	TRUE
2	10504237	Production Engineering	0,9358566 737	4	FALSE	TRUE
3	11890896	Production Engineering	0,9355367 658	1	TRUE	FALSE
4	22861181	Production Engineering	0,9352205 072	3	FALSE	TRUE
5	11522068	Production Engineering	0,9340292 108	5	FALSE	TRUE
1	28290448	Public Relations Officer	0,9476316 491	1	TRUE	TRUE
2	70750649	Public Relations Officer	0,9467867 469	2	TRUE	FALSE
3	22861181	Public Relations Officer	0,9404439 705	4	FALSE	FALSE
4	13115648	Public Relations Officer	0,9393615 415	5	FALSE	FALSE
5	22732234	Public Relations Officer	0,9386659 303	3	TRUE	TRUE

1	26888302	Quality Control Supervisor - Corn Commodity	0,9380053 431	2	FALSE	FALSE
2	22861181	Quality Control Supervisor - Corn Commodity	0,9358159 329	4	FALSE	TRUE
3	20905088	Quality Control Supervisor - Corn Commodity	0,9323824 831	1	FALSE	TRUE
4	21629057	Quality Control Supervisor - Corn Commodity	0,9320897 127	3	FALSE	FALSE
5	26932091	Quality Control Supervisor - Corn Commodity	0,9292874 126	5	FALSE	TRUE
1	26932091	Regional Sales Manager	0,9526580 458	3	TRUE	TRUE
2	28867567	Regional Sales Manager	0,9380075 485	4	TRUE	TRUE
3	29306433	Regional Sales Manager	0,9375366 445	5	TRUE	FALSE
4	27715131	Regional Sales Manager	0,9370184 881	1	TRUE	TRUE
5	28051330	Regional Sales Manager	0,9363337 906	2	TRUE	TRUE

1	22861181	Spare part Admin	0,9482737 396	1	FALSE	FALSE
2	23917826	Spare part Admin	0,9471226 141	3	FALSE	TRUE
3	26932091	Spare part Admin	0,9467146 582	2	FALSE	FALSE
4	16378091	Spare part Admin	0,9466934 216	4	FALSE	TRUE
5	10464113	Spare part Admin	0,9449295 182	5	FALSE	TRUE
1	28772892	Teachers	0,9642666 724	5	TRUE	TRUE
2	15850434	Teachers	0,9575843 692	2	TRUE	TRUE
3	58105060	Teachers	0,9521600 9	4	TRUE	FALSE
4	37220856	Teachers	0,9496124 554	3	TRUE	FALSE
5	48547319	Teachers	0,9471607 207	1	TRUE	FALSE
1	26932091	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,9621574 749	1	TRUE	TRUE
2	11289482	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,9439541 207	3	TRUE	TRUE
3	30862904	Unmanaged Merchant Engagement Senior Associate,	0,9380125 051	4	FALSE	FALSE

		BPO Field Sales				
4	13964744	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,9332395 557	2	FALSE	FALSE
5	24727739	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,9331051 675	5	FALSE	FALSE

#### C.4 Dengan Bobot - TF-IDF dan Improved Sqrt-Cosine Similarity

Rank	Resume ID	Position	Similarity Score	Rank Expert	Relevance	Seniority
1	Business Development Executive	0,205913 2303	2	TRUE	FALSE	38688388
2	Business Development Executive	0,202912 8525	1	TRUE	FALSE	31638814
3	Business Development Executive	0,192598 8694	3	TRUE	FALSE	47067533
4	Business Development Executive	0,187834 9723	5	FALSE	TRUE	26932091
5	Business Development Executive	0,183779 5536	4	TRUE	FALSE	17132168

1	CLUB GENERAL MANAGER	0,204517 8615	5	FALSE	FALSE	26932091
2	CLUB GENERAL MANAGER	0,191228 4503	1	TRUE	TRUE	17818707
3	CLUB GENERAL MANAGER	0,185466 2893	3	TRUE	TRUE	18171955
4	CLUB GENERAL MANAGER	0,181416 6723	4	FALSE	FALSE	15535920
5	CLUB GENERAL MANAGER	0,180679 3312	2	TRUE	TRUE	31761591
1	Construction Supervisor	0,222458 9589	1	TRUE	FALSE	39027764
2	Construction Supervisor	0,219571 2303	4	TRUE	TRUE	12839152
3	Construction Supervisor	0,205773 4896	3	TRUE	FALSE	27246366
4	Construction Supervisor	0,200795 8072	2	TRUE	TRUE	56525735
5	Construction Supervisor	0,196674 0695	5	FALSE	FALSE	26932091
1	Creative Director / Manager	0,177042 6406	2	FALSE	TRUE	68781345
2	Creative Director / Manager	0,152407 9329	3	TRUE	FALSE	18460045
3	Creative Director / Manager	0,149001 695	4	FALSE	FALSE	13964744
4	Creative Director / Manager	0,138984 5056	5	FALSE	FALSE	17781039

5	Creative Director / Manager	0,134054 7236	1	TRUE	TRUE	30864828
1	Digital and Social Media Executive	0,170029 822	5	FALSE	TRUE	22754014
2	Digital and Social Media Executive	0,161378 88	3	TRUE	FALSE	15479281
3	Digital and Social Media Executive	0,147463 8205	1	TRUE	TRUE	16620172
4	Digital and Social Media Executive	0,144642 1125	4	TRUE	FALSE	16536141
5	Digital and Social Media Executive	0,144450 2095	2	TRUE	TRUE	75329822
1	Digital Banking Officer	0,200769 5753	5	FALSE	TRUE	26932091
2	Digital Banking Officer	0,181778 1123	1	TRUE	FALSE	98965485
3	Digital Banking Officer	0,159948 4431	4	FALSE	FALSE	14937492
4	Digital Banking Officer	0,157637 613	3	TRUE	FALSE	27080812
5	Digital Banking Officer	0,154671 2682	2	TRUE	FALSE	29406313
1	Executive Chef	0,260834 3156	4	TRUE	TRUE	29775391
2	Executive Chef	0,260057 9128	3	TRUE	TRUE	20321582

3	Executive Chef	0,249323 4669	1	TRUE	TRUE	34252537
4	Executive Chef	0,246398 796	5	TRUE	TRUE	25924968
5	Executive Chef	0,246316 0122	2	TRUE	TRUE	25128608
1	Finance Executive / Accountant	0,244852 904	5	TRUE	FALSE	37370455
2	Finance Executive / Accountant	0,238935 8751	4	TRUE	FALSE	25846894
3	Finance Executive / Accountant	0,229003 7916	3	TRUE	FALSE	21338490
4	Finance Executive / Accountant	0,218988 459	2	TRUE	FALSE	24670867
5	Finance Executive / Accountant	0,212373 8693	1	TRUE	TRUE	23387174
1	Finance Officer (Jr/Sr.)	0,223954 8861	2	TRUE	TRUE	23734441
2	Finance Officer (Jr/Sr.)	0,216914 026	3	TRUE	TRUE	29999135
3	Finance Officer (Jr/Sr.)	0,216765 7361	5	TRUE	FALSE	24670867
4	Finance Officer (Jr/Sr.)	0,215849 502	1	TRUE	TRUE	28298773
5	Finance Officer (Jr/Sr.)	0,204156 394	4	TRUE	TRUE	53640713
1	Financial Consolidation Consultant	0,210191 4441	5	FALSE	FALSE	68781345
2	Financial Consolidation	0,194978 4098	1	TRUE	TRUE	95792386

	on Consultant					
3	Financial Consolidati on Consultant	0,190158 5017	3	TRUE	TRUE	70541112
4	Financial Consolidati on Consultant	0,189511 6438	2	TRUE	TRUE	38946032
5	Financial Consolidati on Consultant	0,183903 2081	4	TRUE	FALSE	19234823
1	Graphics Designer	0,279561 5313	2	TRUE	TRUE	18460045
2	Graphics Designer	0,222008 2185	3	TRUE	FALSE	18354623
3	Graphics Designer	0,198927 8803	1	TRUE	TRUE	26676567
4	Graphics Designer	0,195168 6831	4	TRUE	FALSE	16893572
5	Graphics Designer	0,191999 01	5	FALSE	TRUE	22754014
1	HR Specialist	0,266403 927	2	TRUE	TRUE	24508725
2	HR Specialist	0,263968 2699	1	TRUE	TRUE	30862904
3	HR Specialist	0,234010 6126	4	FALSE	TRUE	16877897
4	HR Specialist	0,227996 3281	3	FALSE	TRUE	24184357
5	HR Specialist	0,227032 5052	5	FALSE	TRUE	26932091
1	INFORMATI ON & TECHNOLO GY STAFF	0,221170 9312	4	FALSE	FALSE	17983957
2	INFORMATI ON &	0,220545 8455	1	TRUE	TRUE	39413067

	TECHNOLOGY STAFF					
3	INFORMATION & TECHNOLOGY STAFF	0,215223 6294	3	FALSE	FALSE	17570634
4	INFORMATION & TECHNOLOGY STAFF	0,207111 8727	2	TRUE	FALSE	21283365
5	INFORMATION & TECHNOLOGY STAFF	0,202252 2526	5	FALSE	FALSE	38897568
1	Junior Associate Lawyer	0,124307 7984	4	FALSE	TRUE	29406313
2	Junior Associate Lawyer	0,121568 8418	1	TRUE	TRUE	10332998
3	Junior Associate Lawyer	0,120404 3196	3	FALSE	TRUE	69181350
4	Junior Associate Lawyer	0,117563 5815	5	FALSE	TRUE	23636277
5	Junior Associate Lawyer	0,117548 7301	2	TRUE	TRUE	15100547
1	Junior Designer for Apparel	0,166988 3684	2	TRUE	FALSE	23719943
2	Junior Designer for Apparel	0,164602 35	1	TRUE	TRUE	15746146
3	Junior Designer for Apparel	0,150207 8775	5	FALSE	FALSE	12122372
4	Junior Designer for Apparel	0,144722 2826	4	FALSE	FALSE	26932091

5	Junior Designer for Apparel	0,142571 9056	3	FALSE	FALSE	11722421
1	Manager Aviation Safety, Quality and Security	0,234812 8415	2	FALSE	FALSE	26932091
2	Manager Aviation Safety, Quality and Security	0,210154 9087	1	TRUE	FALSE	28186635
3	Manager Aviation Safety, Quality and Security	0,207404 6317	3	FALSE	FALSE	24589765
4	Manager Aviation Safety, Quality and Security	0,201492 0559	4	FALSE	FALSE	29406313
5	Manager Aviation Safety, Quality and Security	0,194965 5358	5	FALSE	FALSE	11289482
1	Medical Doctor	0,225712 691	1	FALSE	TRUE	16356151
2	Medical Doctor	0,174346 7306	5	FALSE	FALSE	13565152
3	Medical Doctor	0,172704 0952	2	FALSE	TRUE	43994605
4	Medical Doctor	0,161099 4638	3	FALSE	FALSE	24588864
5	Medical Doctor	0,152177 5854	4	FALSE	FALSE	49325370
1	Production Engineering	0,144448 7571	2	TRUE	FALSE	28803888
2	Production Engineering	0,130751 2078	4	FALSE	FALSE	30288581

3	Production Engineering	0,128543 7933	1	TRUE	TRUE	77828437
4	Production Engineering	0,123889 6191	3	TRUE	FALSE	17103000
5	Production Engineering	0,121891 3506	5	FALSE	FALSE	54100393
1	Public Relations Officer	0,200719 3257	1	TRUE	TRUE	21297828
2	Public Relations Officer	0,196597 6361	5	FALSE	TRUE	13129275
3	Public Relations Officer	0,195011 6308	2	TRUE	FALSE	27257013
4	Public Relations Officer	0,185810 0177	4	FALSE	FALSE	31220062
5	Public Relations Officer	0,183693 9263	3	TRUE	FALSE	27000192
1	Quality Control Supervisor - Corn Commodity	0,140370 3197	4	FALSE	TRUE	26932091
2	Quality Control Supervisor - Corn Commodity	0,127219 2623	5	FALSE	FALSE	28186635
3	Quality Control Supervisor - Corn Commodity	0,126608 0495	1	FALSE	TRUE	16723524
4	Quality Control Supervisor - Corn Commodity	0,123159 2765	2	FALSE	FALSE	28628090
5	Quality Control Supervisor -	0,121937 0839	3	FALSE	FALSE	20905088

	Corn Commodity					
1	Regional Sales Manager	0,243053 1858	4	FALSE	TRUE	26932091
2	Regional Sales Manager	0,224709 9765	3	FALSE	TRUE	27080812
3	Regional Sales Manager	0,217841 7617	1	FALSE	TRUE	25038571
4	Regional Sales Manager	0,212831 0977	2	FALSE	FALSE	38688388
5	Regional Sales Manager	0,210324 2631	5	FALSE	TRUE	26919036
1	Spare part Admin	0,155213 3324	5	FALSE	FALSE	16911115
2	Spare part Admin	0,149973 7297	4	FALSE	FALSE	38897568
3	Spare part Admin	0,148060 5907	1	TRUE	TRUE	10189110
4	Spare part Admin	0,147032 3157	2	TRUE	TRUE	20504094
5	Spare part Admin	0,146112 3257	3	FALSE	FALSE	24670867
1	Teachers	0,250745 2502	4	TRUE	TRUE	15850434
2	Teachers	0,213840 2469	3	TRUE	TRUE	20399718
3	Teachers	0,200342 6776	5	TRUE	TRUE	28772892
4	Teachers	0,196635 771	2	TRUE	TRUE	22056333
5	Teachers	0,196100 2177	1	TRUE	TRUE	96547039
1	Unmanaged Merchant Engagement Senior Associate,	0,222177 0964	1	TRUE	TRUE	26932091

	BPO Field Sales					
2	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,1921965525	3	FALSE	FALSE	24589765
3	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,1865806571	4	FALSE	FALSE	68781345
4	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,1634891955	2	TRUE	FALSE	11289482
5	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,1617524074	5	FALSE	TRUE	26919036

### C.5 Dengan Bobot - Word2Vec dan Cosine Similarity

Rank	Resume ID	Position	Similarity Score	Rank Expert	Relevance	Seniority
1	26932091	Business Development Executive	0,9016295473	3	TRUE	FALSE
2	10464113	Business Development Executive	0,8664782882	2	TRUE	TRUE

3	14790629	Business Development Executive	0,8652020 196	4	TRUE	TRUE
4	27715131	Business Development Executive	0,8478853 941	1	TRUE	FALSE
5	16276121	Business Development Executive	0,8414156 304	5	FALSE	TRUE
1	13411858	CLUB GENERAL MANAGER	0,8781417 949	1	TRUE	TRUE
2	26932091	CLUB GENERAL MANAGER	0,8773720 443	5	FALSE	FALSE
3	10464113	CLUB GENERAL MANAGER	0,8592673 114	4	FALSE	TRUE
4	27715131	CLUB GENERAL MANAGER	0,8545579 91	2	TRUE	TRUE
5	34033933	CLUB GENERAL MANAGER	0,8522922 794	3	FALSE	FALSE
1	26932091	Construction Supervisor	0,8621428 311	5	FALSE	FALSE
2	12839152	Construction Supervisor	0,8265170 306	2	TRUE	TRUE
3	24589765	Construction Supervisor	0,8242714 703	3	FALSE	FALSE

4	27246366	Construction Supervisor	0,8220116 446	1	TRUE	TRUE
5	39027764	Construction Supervisor	0,8205659 688	4	TRUE	FALSE
1	34033933	Creative Director / Manager	0,8544749 737	2	FALSE	FALSE
2	23917826	Creative Director / Manager	0,8538641 334	3	FALSE	FALSE
3	295257 15	Creative Director / Manager	0,8505293 826	5	FALSE	FALSE
4	28471099	Creative Director / Manager	0,8454253 674	1	FALSE	TRUE
5	13115648	Creative Director / Manager	0,8417996 705	4	FALSE	TRUE
1	18905648	Digital and Social Media Executive	0,8844326 053	5	FALSE	TRUE
2	16276121	Digital and Social Media Executive	0,8792089 298	4	FALSE	TRUE
3	18354623	Digital and Social Media Executive	0,8678290 546	2	TRUE	FALSE
4	22754014	Digital and Social Media Executive	0,8591275 096	3	FALSE	TRUE
5	70750649	Digital and Social Media Executive	0,8561320 066	1	TRUE	FALSE

1	26932091	Digital Banking Officer	0,8992728 761	4	FALSE	TRUE
2	29406313	Digital Banking Officer	0,8862656 554	2	FALSE	FALSE
3	11289482	Digital Banking Officer	0,8644934 257	3	TRUE	FALSE
4	16276121	Digital Banking Officer	0,8610806 125	1	TRUE	TRUE
5	10464113	Digital Banking Officer	0,8505692 522	5	FALSE	TRUE
1	35579812	Executive Chef	0,8859671 312	5	TRUE	TRUE
2	21060367	Executive Chef	0,8854268 859	4	TRUE	TRUE
3	20321582	Executive Chef	0,8789194 787	2	TRUE	TRUE
4	29775391	Executive Chef	0,8756774 834	1	TRUE	TRUE
5	34252537	Executive Chef	0,8704000 922	3	TRUE	TRUE
1	20393721	Finance Executive / Accountant	0,8825229 636	3	TRUE	TRUE
2	23636277	Finance Executive / Accountant	0,8781847 954	2	TRUE	TRUE
3	70541112	Finance Executive / Accountant	0,8773093 777	1	TRUE	TRUE
4	28522529	Finance Executive /	0,8708689 809	4	TRUE	FALSE

		Accountant				
5	11289482	Finance Executive / Accountant	0,8649438 109	5	FALSE	FALSE
1	34198885	Finance Officer ( Jr/Sr.)	0,8998344 839	1	TRUE	TRUE
2	70541112	Finance Officer ( Jr/Sr.)	0,8726109 783	3	TRUE	TRUE
3	28522529	Finance Officer ( Jr/Sr.)	0,8644792 199	4	TRUE	FALSE
4	20393721	Finance Officer ( Jr/Sr.)	0,8589066 515	2	TRUE	TRUE
5	25497147	Finance Officer ( Jr/Sr.)	0,8502537 529	5	FALSE	TRUE
1	70541112	Financial Consolidation Consultant	0,8643537 442	1	TRUE	TRUE
2	18365443	Financial Consolidation Consultant	0,8542422 007	2	FALSE	TRUE
3	269320 91	Financial Consolidation Consultant	0,8518766 099	3	FALSE	TRUE
4	26695839	Financial Consolidation Consultant	0,8398145 871	4	FALSE	FALSE
5	16877897	Financial Consolida	0,8222200 423	5	FALSE	TRUE

		tion Consultan t				
1	18460045	Graphics Designer	0,8983531 892	1	TRUE	TRUE
2	18354623	Graphics Designer	0,8972359 945	2	TRUE	TRUE
3	33893326	Graphics Designer	0,8583144 695	3	TRUE	TRUE
4	22754014	Graphics Designer	0,8526434 749	5	FALSE	FALSE
5	16276121	Graphics Designer	0,8444623 484	4	TRUE	FALSE
1	30862904	HR Specialist	0,9084293 384	1	TRUE	TRUE
2	16877897	HR Specialist	0,9057623 595	2	FALSE	TRUE
3	26932091	HR Specialist	0,9027255 476	5	FALSE	FALSE
4	29134372	HR Specialist	0,8710555 988	3	FALSE	TRUE
5	11289482	HR Specialist	0,8705781 315	4	FALSE	FALSE
1	26932091	INFORMA TION & TECHNOL OGY STAFF	0,8806972 802	5	FALSE	FALSE
2	28471099	INFORMA TION & TECHNOL OGY STAFF	0,8790658 832	3	TRUE	FALSE
3	10839851	INFORMA TION & TECHNOL OGY STAFF	0,8555776 477	1	TRUE	TRUE
4	28672970	INFORMA TION & TECHNOL	0,8508172 572	4	TRUE	FALSE

		OGY STAFF				
5	26341987	INFORMATION & TECHNOLOGY STAFF	0,8465297 371	2	TRUE	FALSE
1	26330995	Junior Associate Lawyer	0,8434696 848	3	FALSE	FALSE
2	24589765	Junior Associate Lawyer	0,8409152 248	2	FALSE	FALSE
3	11289482	Junior Associate Lawyer	0,8380842 897	4	FALSE	FALSE
4	28871170	Junior Associate Lawyer	0,8340103 447	1	TRUE	FALSE
5	81508860	Junior Associate Lawyer	0,8320689 852	5	FALSE	FALSE
1	26932091	Junior Designer for Apparel	0,8670108 12	3	FALSE	FALSE
2	28471099	Junior Designer for Apparel	0,8221067 895	2	FALSE	FALSE
3	19195747	Junior Designer for Apparel	0,8096644 228	1	FALSE	FALSE
4	27715131	Junior Designer for Apparel	0,8087251 828	4	FALSE	FALSE
5	76196367	Junior Designer for Apparel	0,8084914 625	5	FALSE	FALSE

1	13195436	Manager Aviation Safety, Quality and Security	0,8608590 662	1	TRUE	TRUE
2	24589765	Manager Aviation Safety, Quality and Security	0,8604515 94	2	FALSE	TRUE
3	28186635	Manager Aviation Safety, Quality and Security	0,8446557 224	3	FALSE	FALSE
4	26932091	Manager Aviation Safety, Quality and Security	0,8428457 081	5	FALSE	FALSE
5	16877897	Manager Aviation Safety, Quality and Security	0,8410134 763	4	FALSE	FALSE
1	15958967	Medical Doctor	0,8506042 004	3	FALSE	FALSE
2	14667957	Medical Doctor	0,8470290 78	1	FALSE	FALSE
3	28745844	Medical Doctor	0,8464838 862	4	FALSE	FALSE
4	24588864	Medical Doctor	0,8438134 193	5	FALSE	FALSE
5	96260484	Medical Doctor	0,8381138 295	2	FALSE	TRUE
1	54100393	Production Engineering	0,8348578 215	5	FALSE	FALSE

2	24544244	Production Engineering	0,8070656 359	3	FALSE	FALSE
3	30288581	Production Engineering	0,8064919 114	2	FALSE	FALSE
4	37751611	Production Engineering	0,8005253 474	4	FALSE	FALSE
5	17312146	Production Engineering	0,7976875 544	1	TRUE	FALSE
1	28290448	Public Relations Officer	0,8546104 868	1	TRUE	TRUE
2	22754014	Public Relations Officer	0,8373224 586	3	TRUE	TRUE
3	16276121	Public Relations Officer	0,8306562 04	4	FALSE	TRUE
4	70750649	Public Relations Officer	0,8303484 569	2	TRUE	FALSE
5	22861181	Public Relations Officer	0,8297583 163	5	FALSE	FALSE
1	26932091	Quality Control Supervisor - Corn Commodity	0,8705424 666	4	FALSE	TRUE
2	35651876	Quality Control Supervisor - Corn Commodity	0,8361086 76	2	FALSE	FALSE

3	26070334	Quality Control Supervisor - Corn Commodity	0,8279247 06	1	FALSE	TRUE
4	21060367	Quality Control Supervisor - Corn Commodity	0,8256160 915	5	FALSE	FALSE
5	12011623	Quality Control Supervisor - Corn Commodity	0,8178911 124	3	FALSE	FALSE
1	26932091	Regional Sales Manager	0,9204637 706	3	FALSE	TRUE
2	28867567	Regional Sales Manager	0,8558917	2	TRUE	TRUE
3	18368613	Regional Sales Manager	0,8488336 618	4	TRUE	FALSE
4	27715131	Regional Sales Manager	0,8465175 219	1	FALSE	TRUE
5	14790629	Regional Sales Manager	0,8421503 535	5	FALSE	FALSE
1	26932091	Spare part Admin	0,8987249 136	5	FALSE	TRUE
2	16378091	Spare part Admin	0,8544802 836	1	FALSE	TRUE
3	14790629	Spare part Admin	0,8472287 589	4	FALSE	FALSE
4	23917826	Spare part Admin	0,8446965 705	3	FALSE	TRUE
5	37764298	Spare part Admin	0,8442025 726	2	FALSE	TRUE

1	28772892	Teachers	0,9297527 121	4	TRUE	TRUE
2	15850434	Teachers	0,9152821 211	2	TRUE	TRUE
3	54100393	Teachers	0,9012950 182	5	TRUE	TRUE
4	37220856	Teachers	0,8875992 112	3	FALSE	TRUE
5	20399718	Teachers	0,8841772 63	1	TRUE	TRUE
1	26932091	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,9252712 045	1	TRUE	TRUE
2	11289482	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,8667385 379	3	TRUE	FALSE
3	24589765	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,8507761 608	4	FALSE	FALSE
4	29406313	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,8341917 147	2	TRUE	FALSE
5	16877897	Unmanaged Merchant	0,8320381 194	5	FALSE	FALSE

		Engagement Senior Associate, BPO Field Sales				
--	--	--	--	--	--	--

### C.6 Dengan Bobot - Word2Vec dan *Improved Sqrt-Cosine Similarity*

Rank	Resume ID	Position	Similarity Score	Rank Expert	Relevance	Seniority
1	26932091	Business Development Executive	0,9539668 445	2	TRUE	FALSE
2	14790629	Business Development Executive	0,9506780 856	4	TRUE	FALSE
3	91467795	Business Development Executive	0,9475252 332	1	TRUE	TRUE
4	10464113	Business Development Executive	0,9466634 695	5	FALSE	TRUE
5	27715131		0,9436823 33	3	TRUE	TRUE
1	13411858	CLUB GENERAL MANAGER	0,9527268 773	4	FALSE	TRUE
2	28471099	CLUB GENERAL MANAGER	0,9490812 742	2	TRUE	TRUE
3	26932091	CLUB GENERAL MANAGER	0,9488134 43	1	TRUE	TRUE
4	10464113	CLUB GENERAL	0,9477072 07	5	FALSE	FALSE

		MANAGE R				
5	24727739	CLUB GENERAL MANAGER	0,9474657 096	3	FALSE	FALSE
1	26932091	Construction Supervisor	0,9412927 301	5	FALSE	TRUE
2	39027764	Construction Supervisor	0,9390667 24	3	TRUE	FALSE
3	12839152	Construction Supervisor	0,9371928 374	1	TRUE	TRUE
4	21060367	Construction Supervisor	0,9352513 378	4	FALSE	FALSE
5	27246366	Construction Supervisor	0,9326540 113	2	TRUE	TRUE
1	28471099	Creative Director / Manager	0,9530115 037	4	FALSE	TRUE
2	17781039	Creative Director / Manager	0,9483212 768	5	FALSE	FALSE
3	139647 44	Creative Director / Manager	0,9477864 338	2	FALSE	TRUE
4	24589765	Creative Director / Manager	0,9477007 135	3	FALSE	TRUE
5	81508860	Creative Director / Manager	0,9471895 609	1	TRUE	TRUE
1	18905648	Digital and Social	0,9553223 582	2	TRUE	TRUE

		Media Executive				
2	18354623	Digital and Social Media Executive	0,9542773 245	3	TRUE	TRUE
3	16276121	Digital and Social Media Executive	0,9541360 868	1	FALSE	FALSE
4	34712719	Digital and Social Media Executive	0,9525442 674	5	FALSE	TRUE
5	70750649	Digital and Social Media Executive	0,9519086 079	4	TRUE	FALSE
1	26932091	Digital Banking Officer	0,9611610 445	5	FALSE	FALSE
2	16276121	Digital Banking Officer	0,9536183 744	1	FALSE	TRUE
3	29406313	Digital Banking Officer	0,9528114 795	3	FALSE	TRUE
4	14790629	Digital Banking Officer	0,9514819 96	4	FALSE	TRUE
5	28471099	Digital Banking Officer	0,9496212 248	2	FALSE	FALSE
1	35579812	Executive Chef	0,9591079 737	5	FALSE	TRUE
2	29775391	Executive Chef	0,9519746 137	1	TRUE	TRUE
3	21060367	Executive Chef	0,9517975 364	4	FALSE	TRUE
4	16924102	Executive Chef	0,9497517 649	2	TRUE	TRUE

5	10276858	Executive Chef	0,9491744 926	3	TRUE	FALSE
1	23636277	Finance Executive / Accountant	0,9531494 892	3	TRUE	TRUE
2	20393721	Finance Executive / Accountant	0,9519314 231	4	FALSE	TRUE
3	70541112	Finance Executive / Accountant	0,9514578 621	1	TRUE	TRUE
4	24670867	Finance Executive / Accountant	0,9505766 607	2	TRUE	TRUE
5	24953921	Finance Executive / Accountant	0,9484162 525	5	FALSE	FALSE
1	34198885	Finance Officer ( Jr/Sr.)	0,9618073 275	1	TRUE	TRUE
2	25497147	Finance Officer ( Jr/Sr.)	0,9601359 54	3	FALSE	TRUE
3	20393721	Finance Officer ( Jr/Sr.)	0,9587329 409	2	FALSE	TRUE
4	53640713	Finance Officer ( Jr/Sr.)	0,9536418 681	4	FALSE	TRUE
5	28522529	Finance Officer ( Jr/Sr.)	0,9535884 272	5	FALSE	FALSE
1	70541112	Financial Consolida	0,9419954 621	1	TRUE	TRUE

		tion Consultan t				
2	27330027	Financial Consolida tion Consultan t	0,9392107 855	3	FALSE	FALSE
3	139647 44	Financial Consolida tion Consultan t	0,9359215 062	4	FALSE	TRUE
4	18365443	Financial Consolida tion Consultan t	0,9342627 198	5	FALSE	TRUE
5	15363277	Financial Consolida tion Consultan t	0,9341559 481	2	FALSE	TRUE
1	18354623	Graphics Designer	0,9600432 242	2	TRUE	FALSE
2	18460045	Graphics Designer	0,9534482 538	1	TRUE	TRUE
3	16276121	Graphics Designer	0,9531318 897	5	TRUE	FALSE
4	70750649	Graphics Designer	0,9528303 428	4	TRUE	FALSE
5	22754014	Graphics Designer	0,9518864 807	3	TRUE	FALSE
1	30862904	HR Specialist	0,9600555 505	3	TRUE	FALSE
2	16877897	HR Specialist	0,9569724 647	2	TRUE	FALSE
3	26932091	HR Specialist	0,9502614 406	5	FALSE	FALSE
4	24508725	HR Specialist	0,9461360 144	1	TRUE	TRUE

5	11289482	HR Specialist	0,9453240 462	4	FALSE	FALSE
1	28471099	INFORMATION & TECHNOLOGY STAFF	0,9529981 197	3	FALSE	FALSE
2	16911115	INFORMATION & TECHNOLOGY STAFF	0,9469800 961	2	FALSE	TRUE
3	26932091	INFORMATION & TECHNOLOGY STAFF	0,9466821 773	5	FALSE	FALSE
4	10839851	INFORMATION & TECHNOLOGY STAFF	0,9461330 73	1	TRUE	TRUE
5	10549585	INFORMATION & TECHNOLOGY STAFF	0,9442940 569	4	FALSE	FALSE
1	24589765	Junior Associate Lawyer	0,9489688 544	5	FALSE	FALSE
2	26330995	Junior Associate Lawyer	0,9456341 09	2	FALSE	FALSE
3	27375577	Junior Associate Lawyer	0,9414918 355	4	FALSE	FALSE
4	28471099	Junior Associate Lawyer	0,9398351 781	3	FALSE	FALSE
5	11289482	Junior Associate Lawyer	0,9392279 01	1	FALSE	FALSE
1	26932091	Junior Designer	0,9411121 762	2	FALSE	FALSE

		for Apparel				
2	23917826	Junior Designer for Apparel	0,93517944	4	FALSE	FALSE
3	28745844	Junior Designer for Apparel	0,9344192132	5	FALSE	FALSE
4	20553895	Junior Designer for Apparel	0,9317723528	3	FALSE	FALSE
5	70750649	Junior Designer for Apparel	0,9313974978	1	FALSE	FALSE
1	21060367	Manager Aviation Safety, Quality and Security	0,9459566999	5	FALSE	FALSE
2	13195436	Manager Aviation Safety, Quality and Security	0,9425360008	1	TRUE	TRUE
3	12333703	Manager Aviation Safety, Quality and Security	0,9420496615	3	FALSE	FALSE
4	29167286	Manager Aviation Safety, Quality and Security	0,9400526723	2	TRUE	TRUE
5	35651876	Manager Aviation Safety,	0,9394215969	4	FALSE	FALSE

		Quality and Security				
1	14667957	Medical Doctor	0,9379945 709	2	FALSE	TRUE
2	28745844	Medical Doctor	0,9371113 42	5	FALSE	TRUE
3	15958967	Medical Doctor	0,9345641 05	1	FALSE	TRUE
4	24588864	Medical Doctor	0,9319105 3	4	FALSE	TRUE
5	96260484	Medical Doctor	0,9310241 31	3	FALSE	TRUE
1	54100393	Production Engineering	0,9440333 022	4	FALSE	TRUE
2	22861181	Production Engineering	0,9393731 832	3	FALSE	TRUE
3	11890896	Production Engineering	0,9386830 222	1	TRUE	FALSE
4	15850434	Production Engineering	0,9380816 317	3	FALSE	TRUE
5	11522068	Production Engineering	0,9367380 258	5	FALSE	TRUE
1	28290448	Public Relations Officer	0,9475246 991	1	TRUE	TRUE
2	70750649	Public Relations Officer	0,9442786 542	2	TRUE	FALSE
3	13115648	Public Relations Officer	0,9423968 156	5	FALSE	FALSE

4	27000192	Public Relations Officer	0,9418473 849	3	TRUE	FALSE
5	22732234	Public Relations Officer	0,9413068 995	4	TRUE	TRUE
1	26888302	Quality Control Supervisor - Corn Commodity	0,9397145 7	3	FALSE	FALSE
2	28628090	Quality Control Supervisor - Corn Commodity	0,9374573 193	1	FALSE	FALSE
3	22861181	Quality Control Supervisor - Corn Commodity	0,9352906 635	4	FALSE	TRUE
4	20905088	Quality Control Supervisor - Corn Commodity	0,9336073 888	2	FALSE	TRUE
5	21060367	Quality Control Supervisor - Corn Commodity	0,9334563 762	5	FALSE	FALSE
1	26932091	Regional Sales Manager	0,9537964 128	3	TRUE	TRUE
2	27715131	Regional Sales Manager	0,9388897 552	1	TRUE	TRUE
3	14790629	Regional Sales Manager	0,9382742 131	4	FALSE	FALSE

4	23917826	Regional Sales Manager	0,9381668 276	5	FALSE	TRUE
5	14070138	Regional Sales Manager	0,9381297 179	2	TRUE	TRUE
1	14790629	Spare part Admin	0,9541040 064	5	FALSE	FALSE
2	26932091	Spare part Admin	0,9529997 728	2	FALSE	FALSE
3	16378091	Spare part Admin	0,9487359 951	3	FALSE	TRUE
4	28745844	Spare part Admin	0,9476989 518	1	TRUE	TRUE
5	23917826	Spare part Admin	0,9474161 594	4	FALSE	TRUE
1	28772892	Teachers	0,9650821 12	5	TRUE	TRUE
2	15850434	Teachers	0,9614404 383	2	TRUE	TRUE
3	58105060	Teachers	0,9525103 733	4	TRUE	FALSE
4	20399718	Teachers	0,9482750 764	1	TRUE	TRUE
5	46055835	Teachers	0,9476272 471	3	FALSE	FALSE
1	26932091	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,9595202 941	1	TRUE	TRUE
2	29406313	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,9428494 614	2	TRUE	TRUE

3	11289482	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,9400499 924	4	TRUE	TRUE
4	30862904	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,9388757 821	5	FALSE	FALSE
5	24589765	Unmanaged Merchant Engagement Senior Associate, BPO Field Sales	0,9388679 936	3	TRUE	TRUE

## LAMPIRAN D GRAFIK GARIS TIGA PARAMETER SETIAP KUALIFIKASI LOWONGAN KERJA

Nilai Korelasi untuk Business Development Executive: [0.0999999999999998, -0.6000000000000001, -0.19999999999996, 0.8, 0.3000000000000004, 0.3000000000000004]

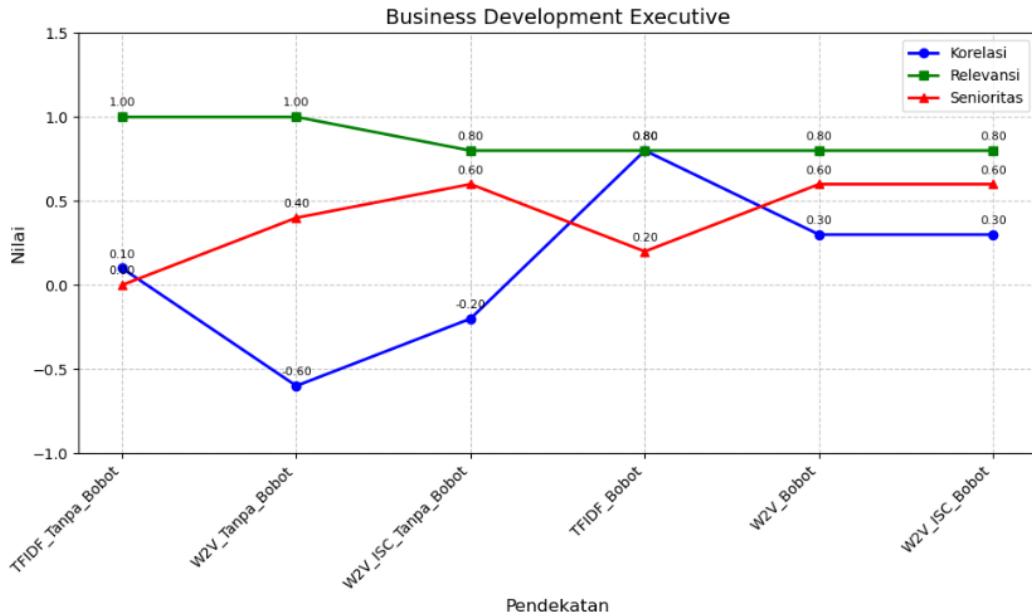
Percentase Relevance untuk Business Development Executive: [1.0, 1.0, 0.8, 0.8, 0.8, 0.8]

Percentase Senioritas untuk Business Development Executive: [0.0, 0.4, 0.6, 0.2, 0.6, 0.6]

Rata-Rata Korelasi: 0.1167, Nilai Tertinggi Korelasi: 0.8000

Rata-Rata Relevansi: 0.8667, Nilai Tertinggi Relevansi: 1.0000

Rata-Rata Senioritas: 0.4000, Nilai Tertinggi Senioritas: 0.6000



Nilai Korelasi untuk CLUB GENERAL MANAGER: [-0.7, -1.0, 0.0, -0.3000000000000004, 0.0999999999999998, 0.0999999999999998]

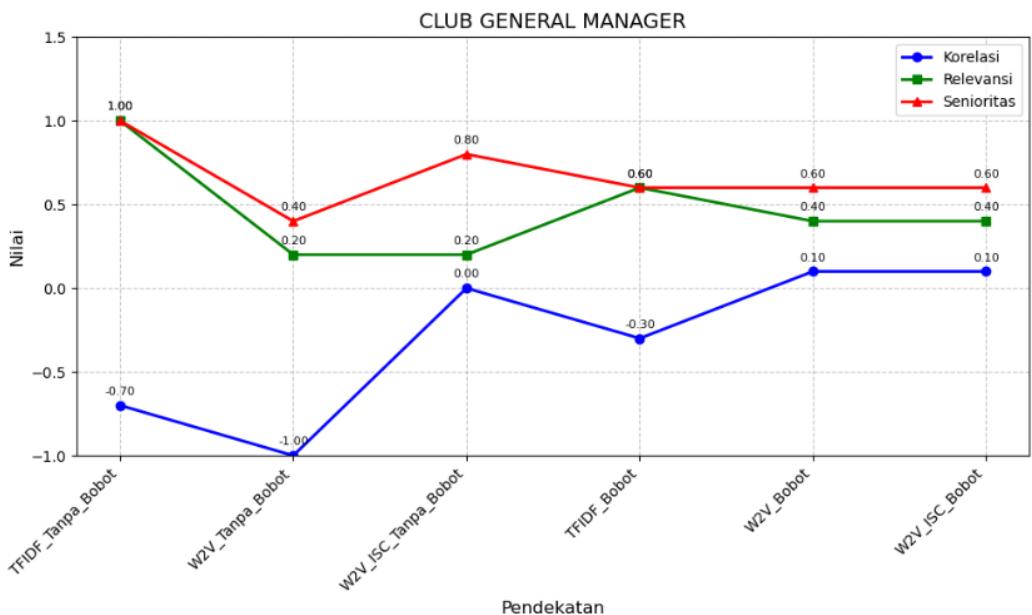
Percentase Relevance untuk CLUB GENERAL MANAGER: [1.0, 0.2, 0.2, 0.6, 0.4, 0.4]

Percentase Senioritas untuk CLUB GENERAL MANAGER: [1.0, 0.4, 0.8, 0.6, 0.6, 0.6]

Rata-Rata Korelasi: -0.3000, Nilai Tertinggi Korelasi: 0.1000

Rata-Rata Relevansi: 0.4667, Nilai Tertinggi Relevansi: 1.0000

Rata-Rata Senioritas: 0.6667, Nilai Tertinggi Senioritas: 1.0000



Nilai Korelasi untuk Construction Supervisor: [0.9, 0.0999999999999998, 0.0999999999999998, 0.6, -0.3000000000000004, -0.5]

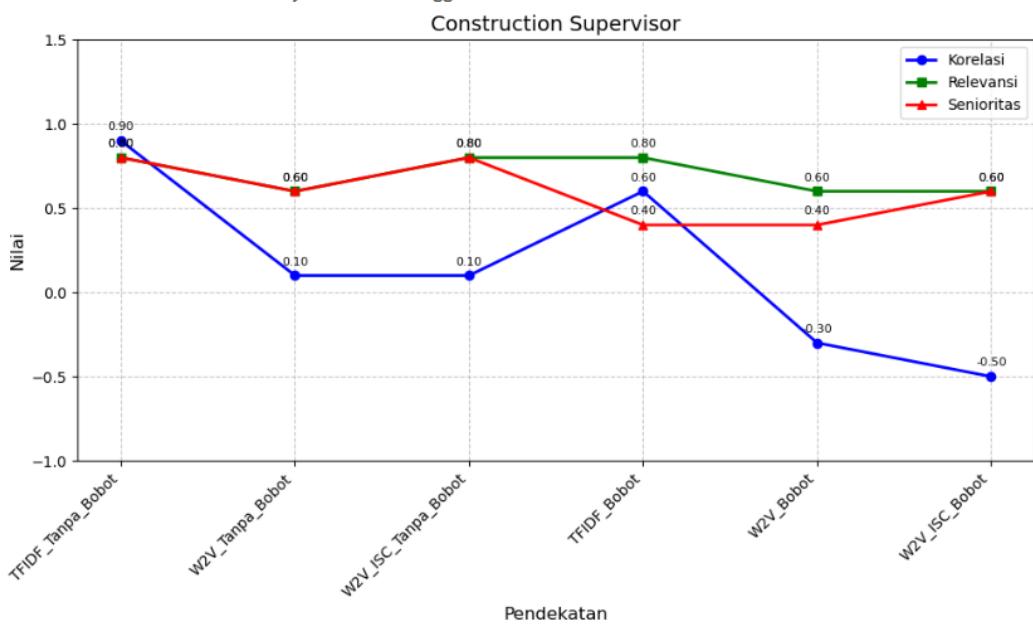
Percentase Relevance untuk Construction Supervisor: [0.8, 0.6, 0.8, 0.8, 0.6, 0.6]

Percentase Senioritas untuk Construction Supervisor: [0.8, 0.6, 0.8, 0.4, 0.4, 0.6]

Rata-Rata Korelasi: 0.1500, Nilai Tertinggi Korelasi: 0.9000

Rata-Rata Relevansi: 0.7000, Nilai Tertinggi Relevansi: 0.8000

Rata-Rata Senioritas: 0.6000, Nilai Tertinggi Senioritas: 0.8000



Nilai Korelasi untuk Creative Director / Manager: [-0.3000000000000004, 0.5, 0.1999999999999996, 0.0, 0.1999999999999996, -0.8]

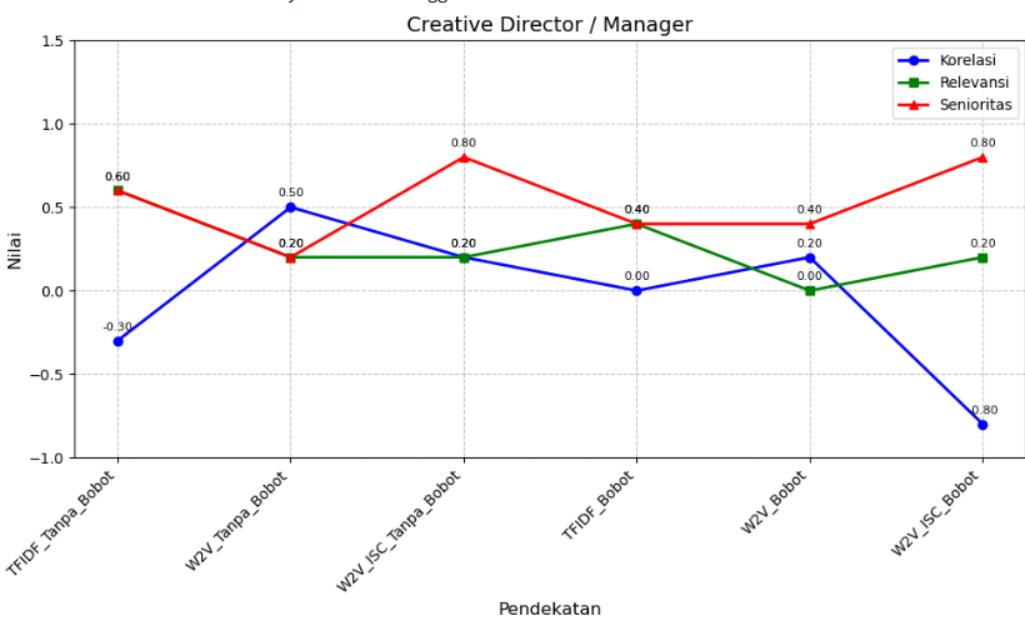
Percentase Relevance untuk Creative Director / Manager: [0.6, 0.2, 0.2, 0.4, 0.0, 0.2]

Percentase Senioritas untuk Creative Director / Manager: [0.6, 0.2, 0.8, 0.4, 0.4, 0.8]

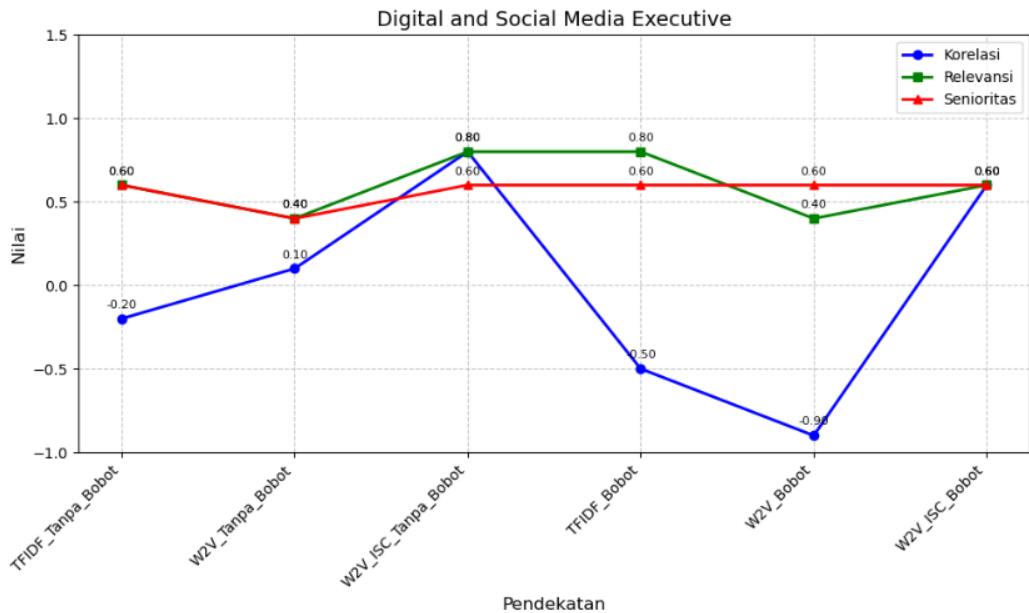
Rata-Rata Korelasi: -0.0333, Nilai Tertinggi Korelasi: 0.5000

Rata-Rata Relevansi: 0.2667, Nilai Tertinggi Relevansi: 0.6000

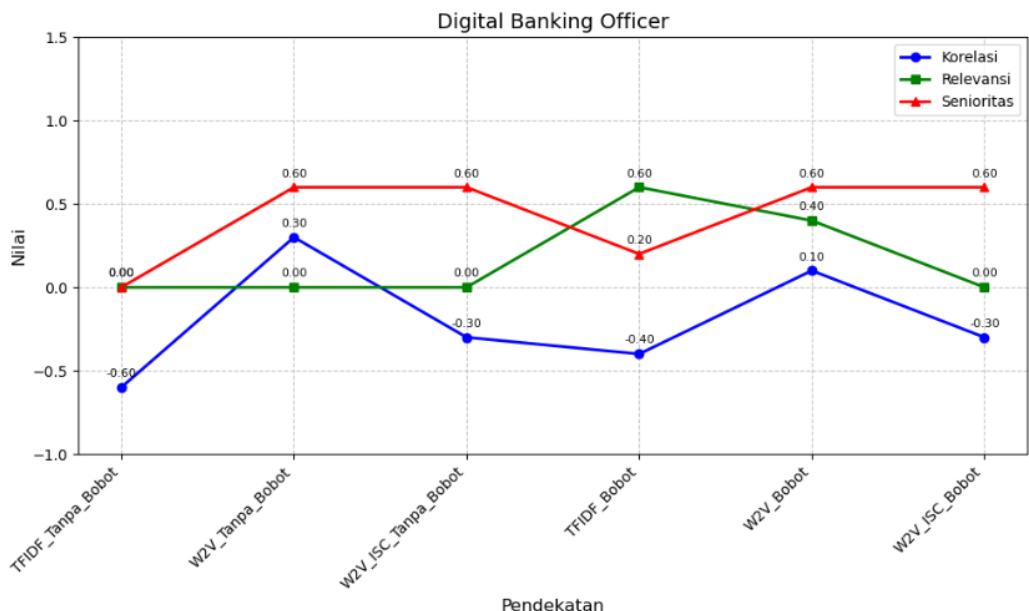
Rata-Rata Senioritas: 0.5333, Nilai Tertinggi Senioritas: 0.8000



Nilai Korelasi untuk Digital and Social Media Executive: [-0.1999999999999996, 0.0999999999999998, 0.8, -0.5, -0.899999999999999, 0.6]  
 Persentase Relevance untuk Digital and Social Media Executive: [0.6, 0.4, 0.8, 0.8, 0.4, 0.6]  
 Persentase Senioritas untuk Digital and Social Media Executive: [0.6, 0.4, 0.6, 0.6, 0.6, 0.6]  
 Rata-Rata Korelasi: -0.0167, Nilai Tertinggi Korelasi: 0.8000  
 Rata-Rata Relevansi: 0.6000, Nilai Tertinggi Relevansi: 0.8000  
 Rata-Rata Senioritas: 0.5667, Nilai Tertinggi Senioritas: 0.6000



Nilai Korelasi untuk Digital Banking Officer: [-0.6000000000000001, 0.3000000000000004, -0.3000000000000004, 0.0000004, -0.399999999999999, 0.0999999999999998, -0.3000000000000004]  
 Persentase Relevance untuk Digital Banking Officer: [0.0, 0.0, 0.0, 0.6, 0.4, 0.0]  
 Persentase Senioritas untuk Digital Banking Officer: [0.0, 0.6, 0.6, 0.2, 0.6, 0.6]  
 Rata-Rata Korelasi: -0.2000, Nilai Tertinggi Korelasi: 0.3000  
 Rata-Rata Relevansi: 0.1667, Nilai Tertinggi Relevansi: 0.6000  
 Rata-Rata Senioritas: 0.4333, Nilai Tertinggi Senioritas: 0.6000



Nilai Korelasi untuk Executive Chef: [0.5, -0.8, -0.7, -0.1999999999999996, -0.7, -0.30000000000000004]

Percentase Relevance untuk Executive Chef: [1.0, 1.0, 0.8, 1.0, 1.0, 0.6]

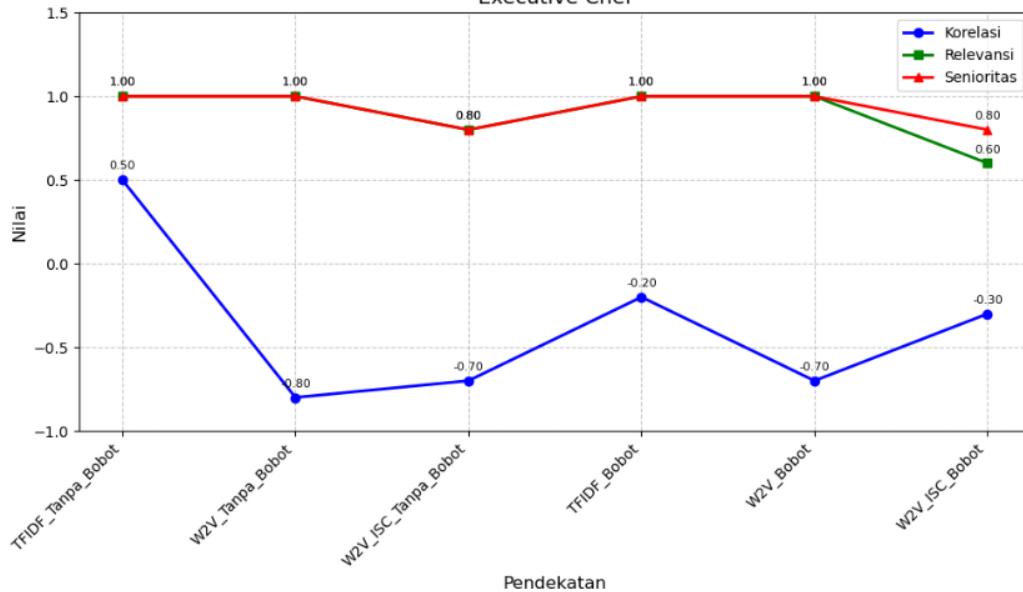
Percentase Senioritas untuk Executive Chef: [1.0, 1.0, 0.8, 1.0, 1.0, 0.8]

Rata-Rata Korelasi: -0.3667, Nilai Tertinggi Korelasi: 0.5000

Rata-Rata Relevansi: 0.9000, Nilai Tertinggi Relevansi: 1.0000

Rata-Rata Senioritas: 0.9333, Nilai Tertinggi Senioritas: 1.0000

**Executive Chef**



Nilai Korelasi untuk Finance Executive / Accountant: [-0.899999999999999, 0.5, 0.7, -1.0, 0.6, 0.1999999999999996]

Percentase Relevance untuk Finance Executive / Accountant: [1.0, 0.6, 0.4, 1.0, 0.8, 0.6]

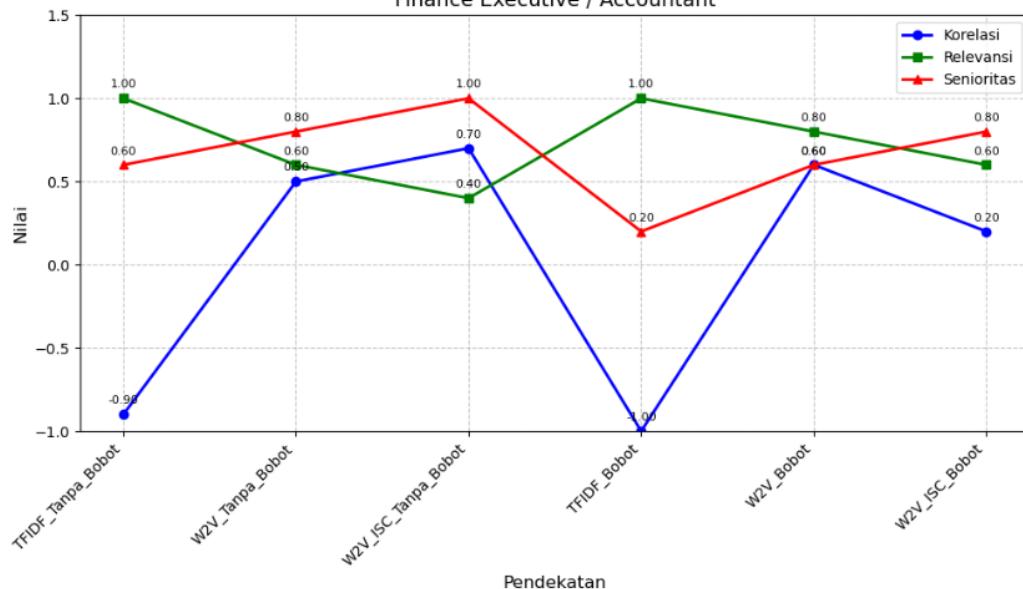
Percentase Senioritas untuk Finance Executive / Accountant: [0.6, 0.8, 1.0, 0.2, 0.6, 0.8]

Rata-Rata Korelasi: 0.0167, Nilai Tertinggi Korelasi: 0.7000

Rata-Rata Relevansi: 0.7333, Nilai Tertinggi Relevansi: 1.0000

Rata-Rata Senioritas: 0.6667, Nilai Tertinggi Senioritas: 1.0000

**Finance Executive / Accountant**



Nilai Korelasi untuk Finance Officer ( Jr/Sr.): [0.7, 0.7, -0.1000000000000009, 0.1999999999999996, 0.7, 0.9]

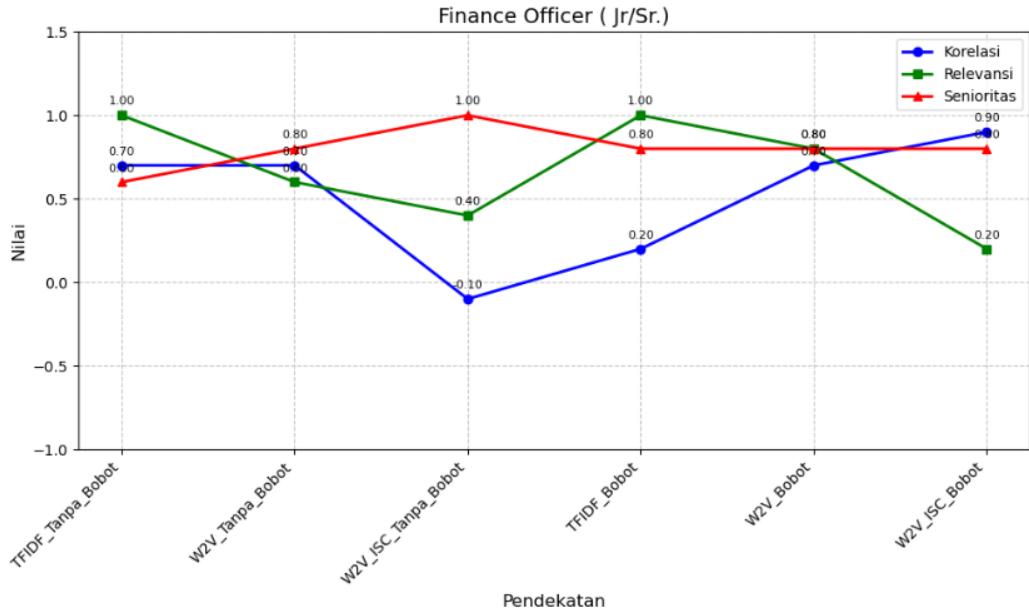
Percentase Relevance untuk Finance Officer ( Jr/Sr.): [1.0, 0.6, 0.4, 1.0, 0.8, 0.2]

Percentase Senioritas untuk Finance Officer ( Jr/Sr.): [0.6, 0.8, 1.0, 0.8, 0.8, 0.8]

Rata-Rata Korelasi: 0.5167, Nilai Tertinggi Korelasi: 0.9000

Rata-Rata Relevansi: 0.6667, Nilai Tertinggi Relevansi: 1.0000

Rata-Rata Senioritas: 0.8000, Nilai Tertinggi Senioritas: 1.0000



Nilai Korelasi untuk Financial Consolidation Consultant: [-0.1000000000000009, 0.7, 0.3000000000000000 04, -0.1000000000000009, 1.0, 0.4]

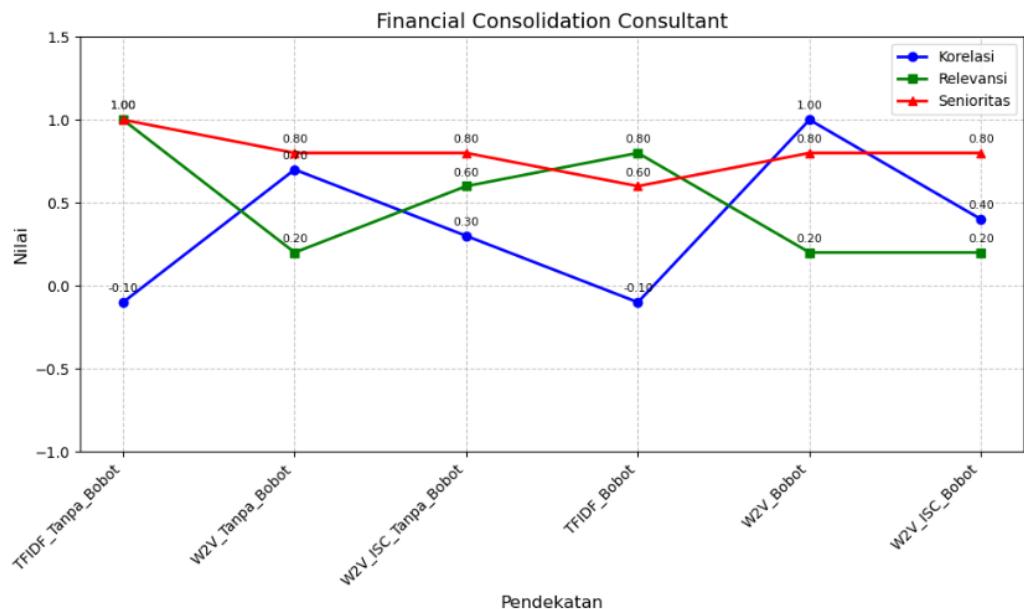
Percentase Relevance untuk Financial Consolidation Consultant: [1.0, 0.2, 0.6, 0.8, 0.2, 0.2]

Percentase Senioritas untuk Financial Consolidation Consultant: [1.0, 0.8, 0.8, 0.6, 0.8, 0.8]

Rata-Rata Korelasi: 0.3667, Nilai Tertinggi Korelasi: 1.0000

Rata-Rata Relevansi: 0.5000, Nilai Tertinggi Relevansi: 1.0000

Rata-Rata Senioritas: 0.8000, Nilai Tertinggi Senioritas: 1.0000



Nilai Korelasi untuk Graphics Designer: [-0.3000000000000004, 0.3000000000000004, 0.5, 0.7, 0.9, 0.5]

Percentase Relevance untuk Graphics Designer: [0.8, 0.6, 1.0, 0.8, 0.8, 1.0]

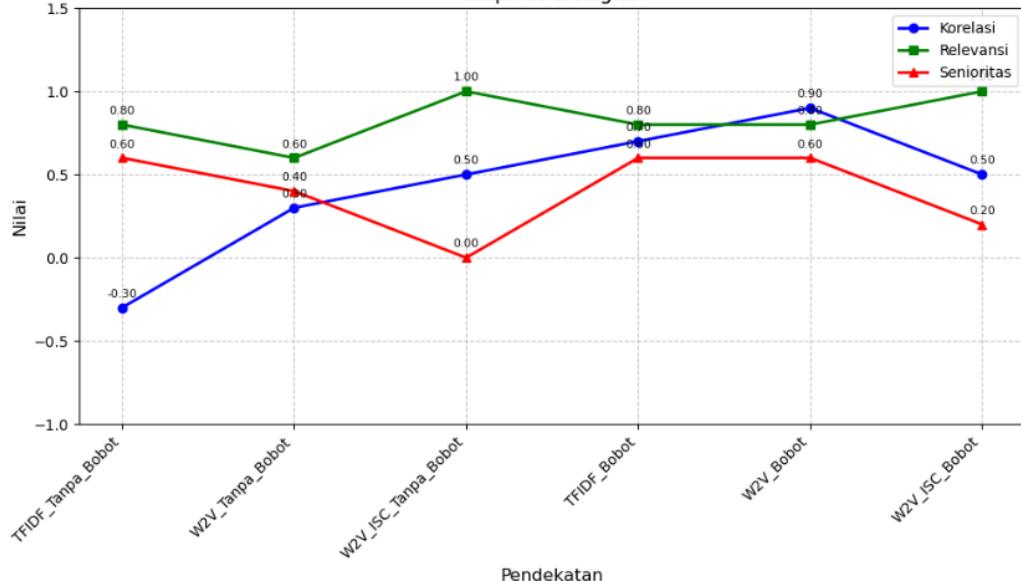
Percentase Senioritas untuk Graphics Designer: [0.6, 0.4, 0.0, 0.6, 0.6, 0.2]

Rata-Rata Korelasi: 0.4333, Nilai Tertinggi Korelasi: 0.9000

Rata-Rata Relevansi: 0.8333, Nilai Tertinggi Relevansi: 1.0000

Rata-Rata Senioritas: 0.4000, Nilai Tertinggi Senioritas: 0.6000

Graphics Designer



Nilai Korelasi untuk HR Specialist: [0.7, 0.1999999999999996, 0.7, 0.8, 0.7, 0.0999999999999998]

Percentase Relevance untuk HR Specialist: [0.4, 0.2, 0.4, 0.4, 0.2, 0.6]

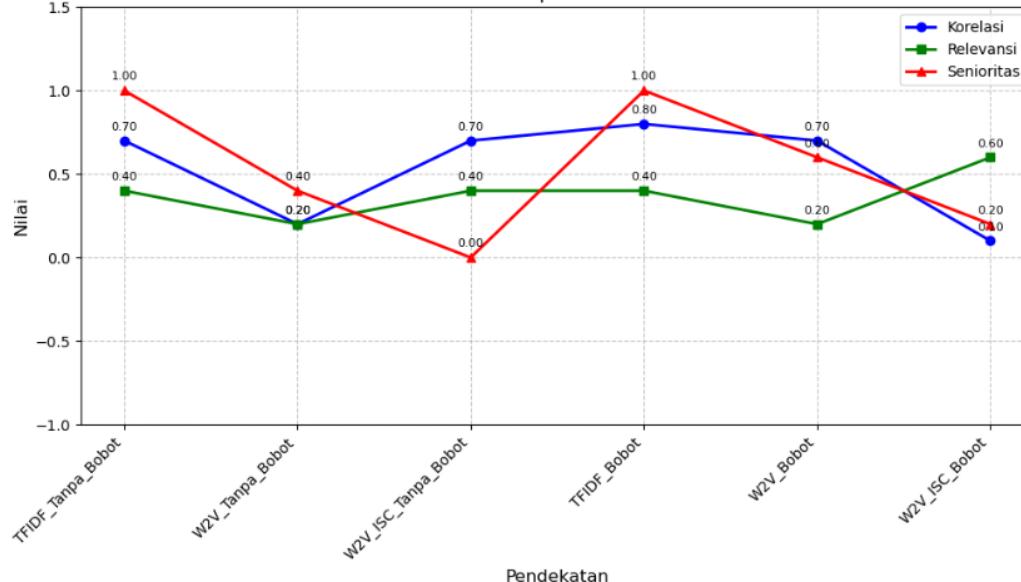
Percentase Senioritas untuk HR Specialist: [1.0, 0.4, 0.0, 1.0, 0.6, 0.2]

Rata-Rata Korelasi: 0.5333, Nilai Tertinggi Korelasi: 0.8000

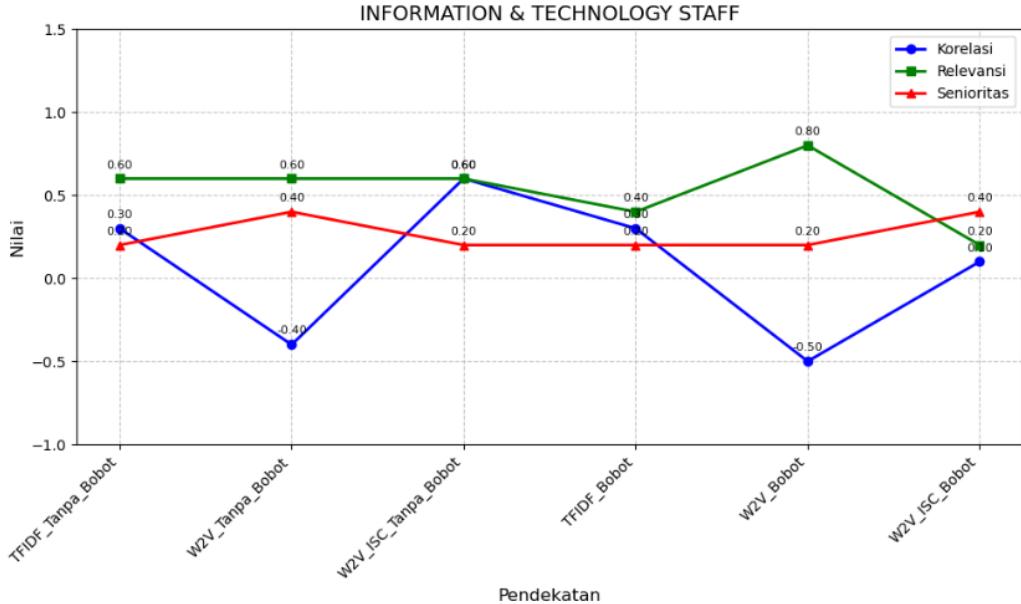
Rata-Rata Relevansi: 0.3667, Nilai Tertinggi Relevansi: 0.6000

Rata-Rata Senioritas: 0.5333, Nilai Tertinggi Senioritas: 1.0000

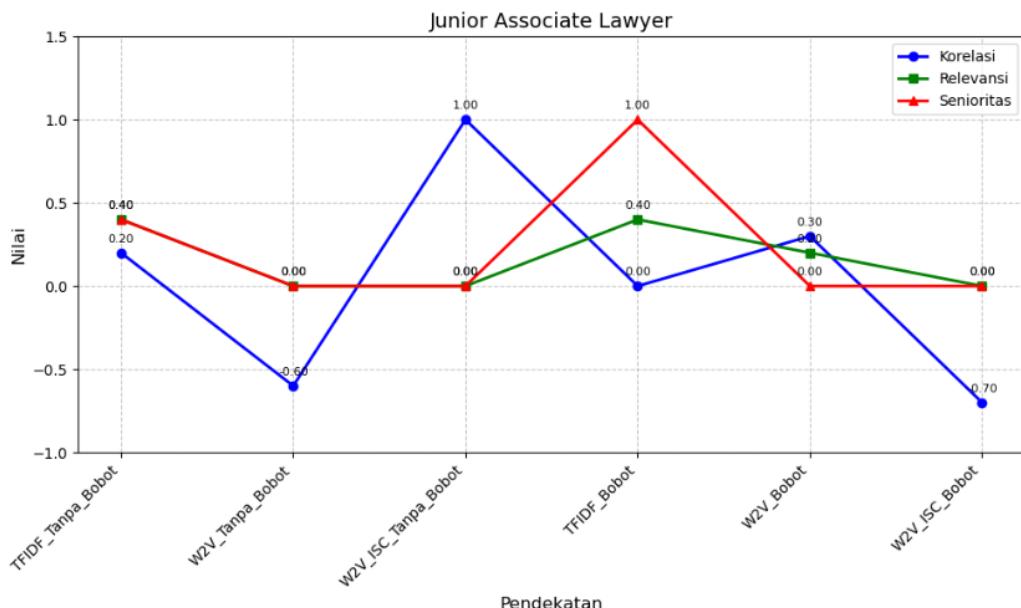
HR Specialist



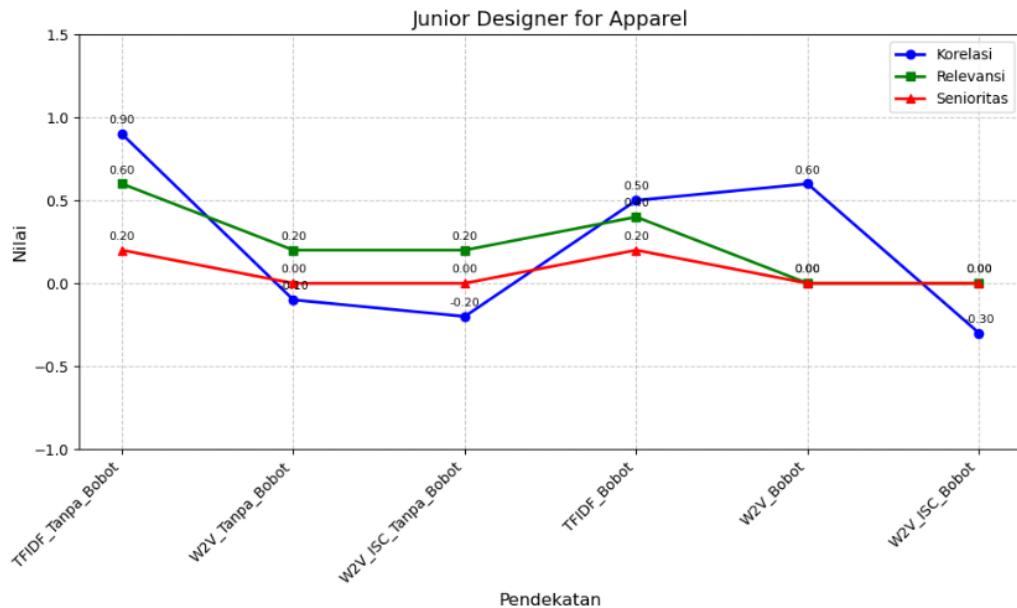
Nilai Korelasi untuk INFORMATION & TECHNOLOGY STAFF: [0.3000000000000004, -0.3999999999999999, 0.6, 0.3000000000000004, -0.5, 0.0999999999999998]  
 Persentase Relevance untuk INFORMATION & TECHNOLOGY STAFF: [0.6, 0.6, 0.6, 0.4, 0.8, 0.2]  
 Persentase Senioritas untuk INFORMATION & TECHNOLOGY STAFF: [0.2, 0.4, 0.2, 0.2, 0.2, 0.4]  
 Rata-Rata Korelasi: 0.0667, Nilai Tertinggi Korelasi: 0.6000  
 Rata-Rata Relevansi: 0.5333, Nilai Tertinggi Relevansi: 0.8000  
 Rata-Rata Senioritas: 0.2667, Nilai Tertinggi Senioritas: 0.4000



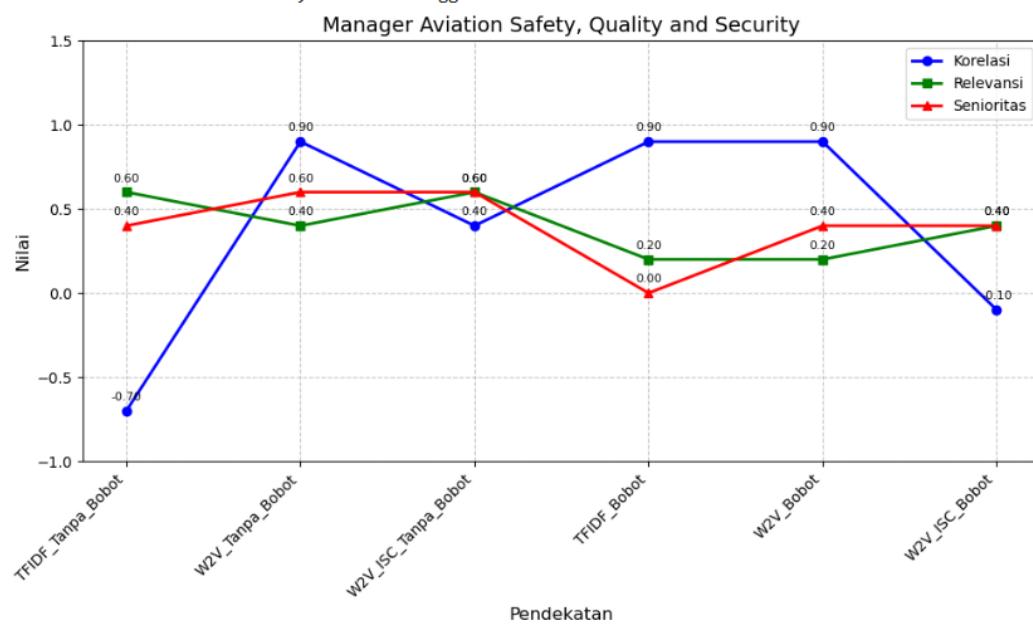
Nilai Korelasi untuk Junior Associate Lawyer: [0.1999999999999996, -0.6000000000000001, 1.0, 0.0, 0.3000000000000004, -0.7]  
 Persentase Relevance untuk Junior Associate Lawyer: [0.4, 0.0, 0.0, 0.4, 0.2, 0.0]  
 Persentase Senioritas untuk Junior Associate Lawyer: [0.4, 0.0, 0.0, 1.0, 0.0, 0.0]  
 Rata-Rata Korelasi: 0.0333, Nilai Tertinggi Korelasi: 1.0000  
 Rata-Rata Relevansi: 0.1667, Nilai Tertinggi Relevansi: 0.4000  
 Rata-Rata Senioritas: 0.2333, Nilai Tertinggi Senioritas: 1.0000



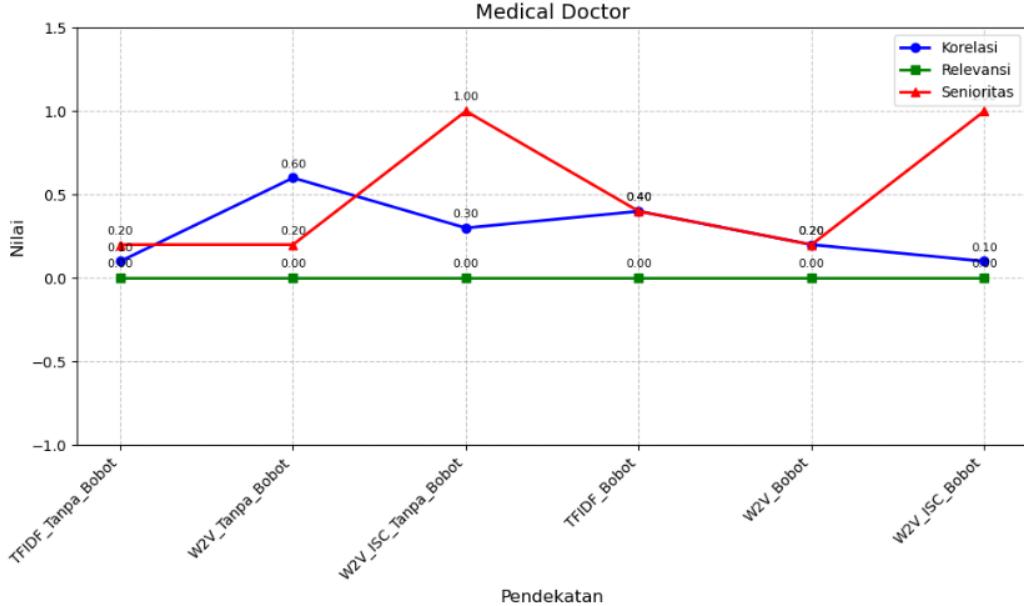
Nilai Korelasi untuk Junior Designer for Apparel: [0.9, -0.1000000000000009, -0.1999999999999996, 0.5, 0.6, -0.3000000000000004]  
 Persentase Relevance untuk Junior Designer for Apparel: [0.6, 0.2, 0.2, 0.4, 0.0, 0.0]  
 Persentase Senioritas untuk Junior Designer for Apparel: [0.2, 0.0, 0.0, 0.2, 0.0, 0.0]  
 Rata-Rata Korelasi: 0.2333, Nilai Tertinggi Korelasi: 0.9000  
 Rata-Rata Relevansi: 0.2333, Nilai Tertinggi Relevansi: 0.6000  
 Rata-Rata Senioritas: 0.0667, Nilai Tertinggi Senioritas: 0.2000



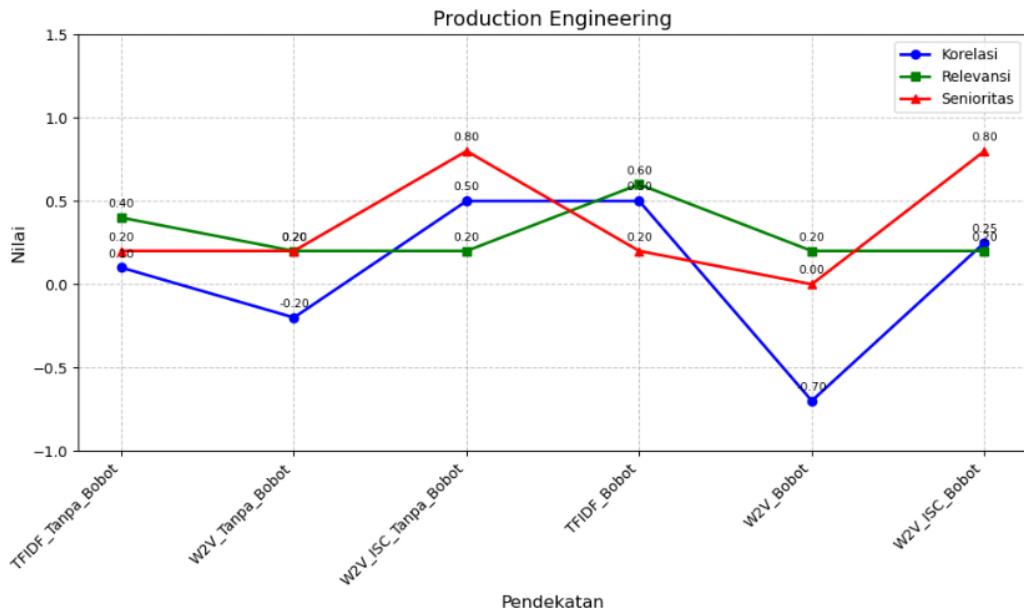
Nilai Korelasi untuk Manager Aviation Safety, Quality and Security: [-0.7, 0.9, 0.4, 0.9, 0.9, -0.100000000009]  
 Persentase Relevance untuk Manager Aviation Safety, Quality and Security: [0.6, 0.4, 0.6, 0.2, 0.2, 0.4]  
 Persentase Senioritas untuk Manager Aviation Safety, Quality and Security: [0.4, 0.6, 0.6, 0.0, 0.4, 0.4]  
 Rata-Rata Korelasi: 0.3833, Nilai Tertinggi Korelasi: 0.9000  
 Rata-Rata Relevansi: 0.4000, Nilai Tertinggi Relevansi: 0.6000  
 Rata-Rata Senioritas: 0.4000, Nilai Tertinggi Senioritas: 0.6000



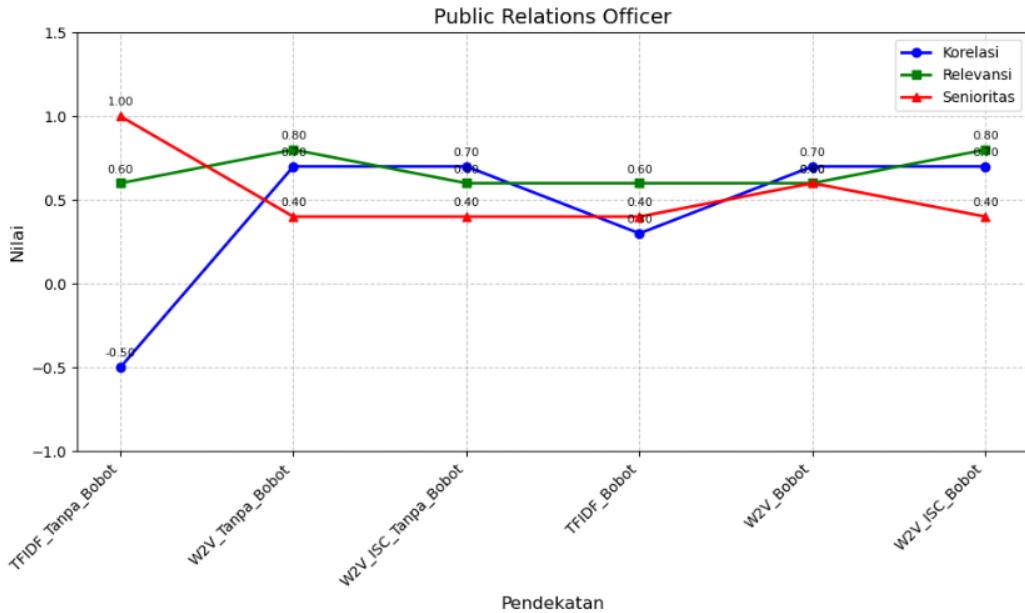
Nilai Korelasi untuk Medical Doctor: [0.0999999999999998, 0.6, 0.3000000000000004, 0.4, 0.1999999999999996, 0.0999999999999998]  
 Persentase Relevance untuk Medical Doctor: [0.0, 0.0, 0.0, 0.0, 0.0, 0.0]  
 Persentase Senioritas untuk Medical Doctor: [0.2, 0.2, 1.0, 0.4, 0.2, 1.0]  
 Rata-Rata Korelasi: 0.2833, Nilai Tertinggi Korelasi: 0.6000  
 Rata-Rata Relevansi: 0.0000, Nilai Tertinggi Relevansi: 0.0000  
 Rata-Rata Senioritas: 0.5000, Nilai Tertinggi Senioritas: 1.0000



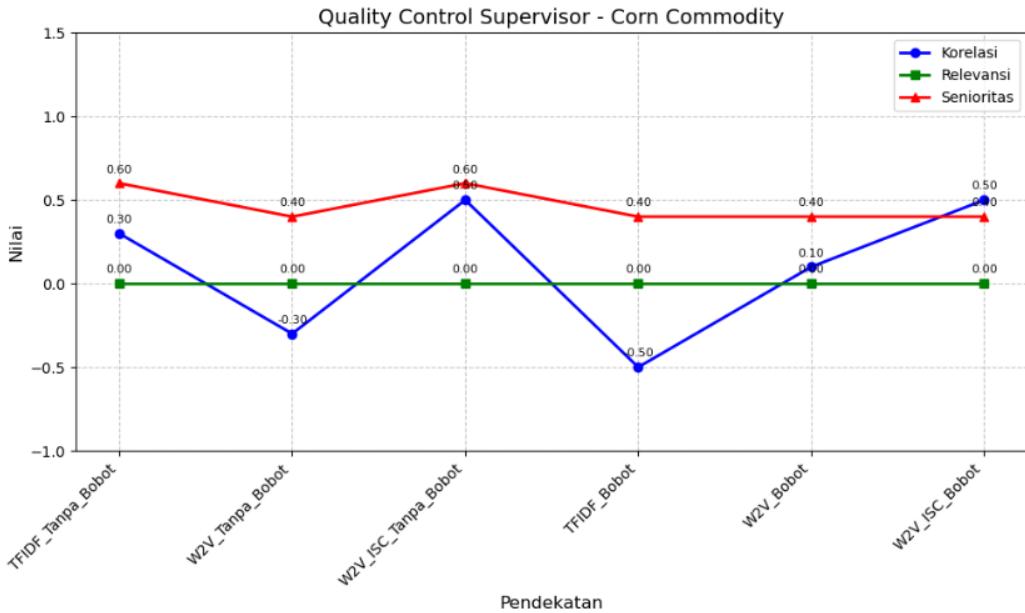
Nilai Korelasi untuk Production Engineering: [0.0999999999999998, -0.1999999999999996, 0.5, 0.5, -0.7, 0.25]  
 Persentase Relevance untuk Production Engineering: [0.4, 0.2, 0.2, 0.6, 0.2, 0.2]  
 Persentase Senioritas untuk Production Engineering: [0.2, 0.2, 0.8, 0.2, 0.0, 0.8]  
 Rata-Rata Korelasi: 0.0750, Nilai Tertinggi Korelasi: 0.5000  
 Rata-Rata Relevansi: 0.3000, Nilai Tertinggi Relevansi: 0.6000  
 Rata-Rata Senioritas: 0.3667, Nilai Tertinggi Senioritas: 0.8000



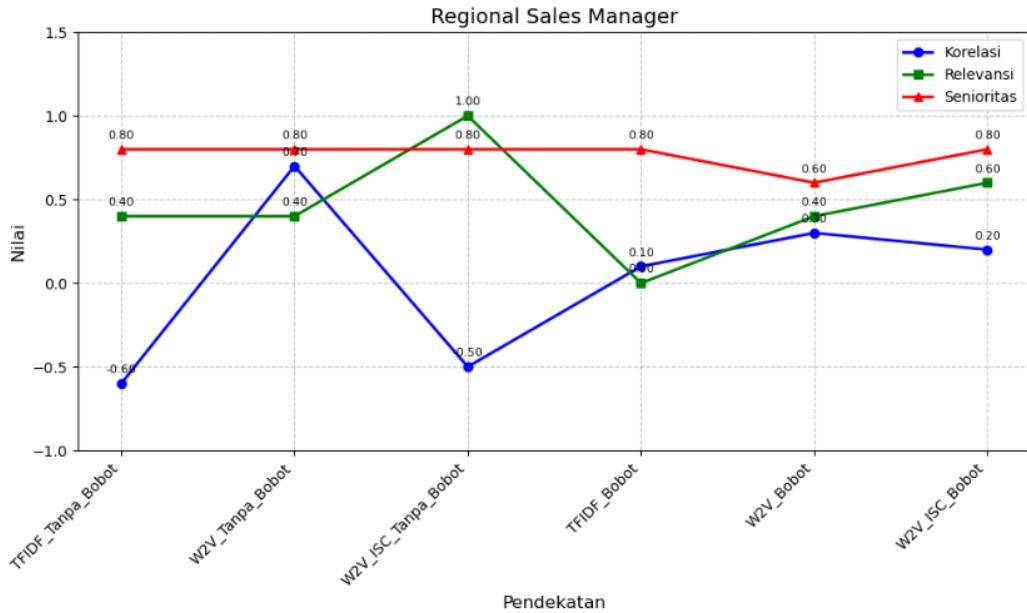
Nilai Korelasi untuk Public Relations Officer: [-0.5, 0.7, 0.7, 0.3000000000000004, 0.7, 0.7]  
 Persentase Relevansi untuk Public Relations Officer: [0.6, 0.8, 0.6, 0.6, 0.6, 0.8]  
 Persentase Senioritas untuk Public Relations Officer: [1.0, 0.4, 0.4, 0.4, 0.6, 0.4]  
 Rata-Rata Korelasi: 0.4333, Nilai Tertinggi Korelasi: 0.7000  
 Rata-Rata Relevansi: 0.6667, Nilai Tertinggi Relevansi: 0.8000  
 Rata-Rata Senioritas: 0.5333, Nilai Tertinggi Senioritas: 1.0000



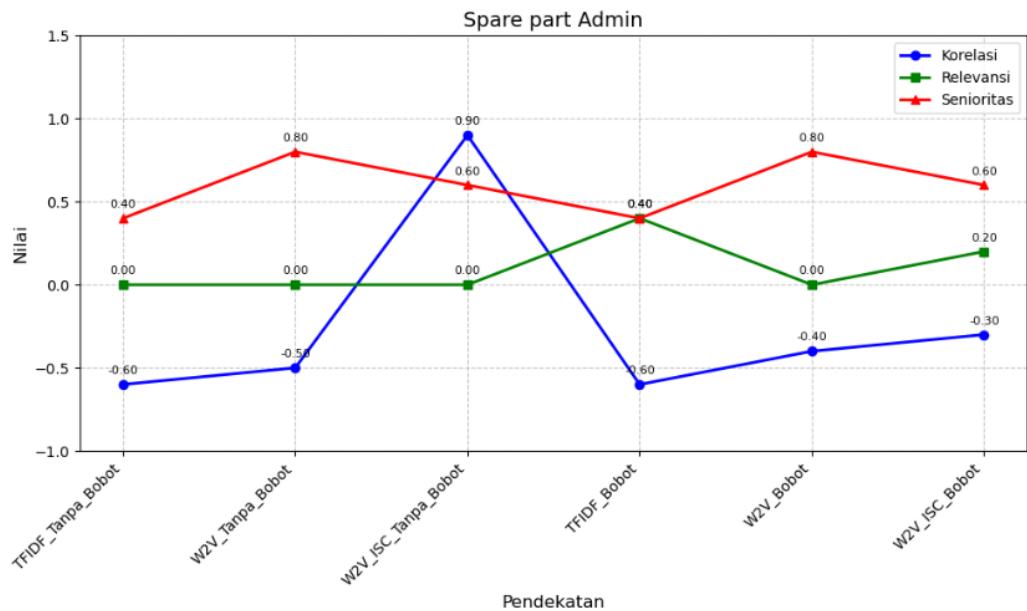
Nilai Korelasi untuk Quality Control Supervisor - Corn Commodity: [0.3000000000000004, -0.3000000000000004, 0.5, -0.5, 0.0999999999999998, 0.5]  
 Persentase Relevansi untuk Quality Control Supervisor - Corn Commodity: [0.0, 0.0, 0.0, 0.0, 0.0, 0.0]  
 Persentase Senioritas untuk Quality Control Supervisor - Corn Commodity: [0.6, 0.4, 0.6, 0.4, 0.4, 0.4]  
 Rata-Rata Korelasi: 0.1000, Nilai Tertinggi Korelasi: 0.5000  
 Rata-Rata Relevansi: 0.0000, Nilai Tertinggi Relevansi: 0.0000  
 Rata-Rata Senioritas: 0.4667, Nilai Tertinggi Senioritas: 0.6000



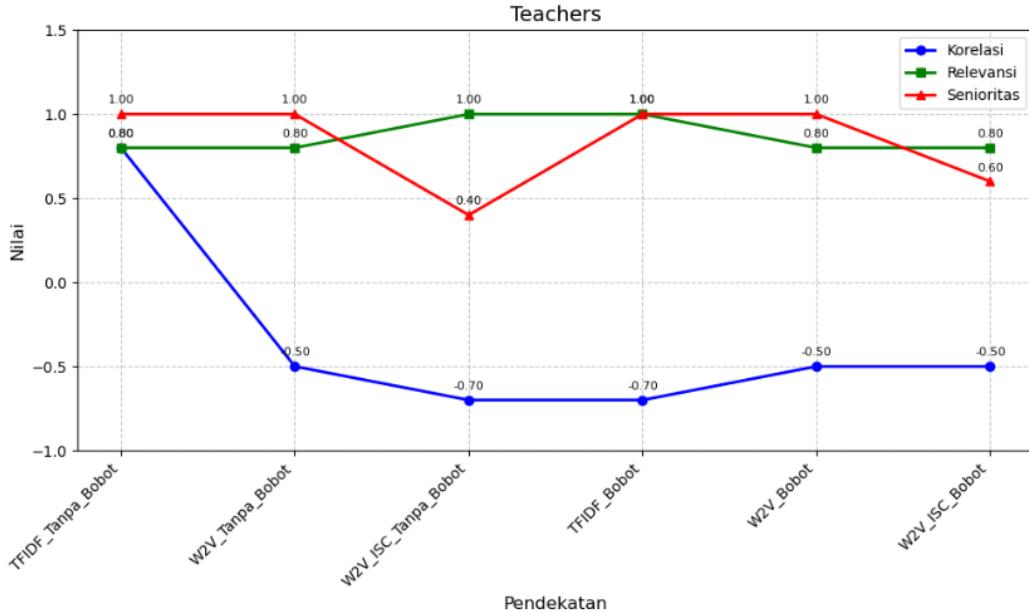
Nilai Korelasi untuk Regional Sales Manager: [-0.6000000000000001, 0.7, -0.5, 0.0999999999999998, 0.3  
 000000000000004, 0.1999999999999996]  
 Persentase Relevance untuk Regional Sales Manager: [0.4, 0.4, 1.0, 0.0, 0.4, 0.6]  
 Persentase Senioritas untuk Regional Sales Manager: [0.8, 0.8, 0.8, 0.8, 0.6, 0.8]  
 Rata-Rata Korelasi: 0.0333, Nilai Tertinggi Korelasi: 0.7000  
 Rata-Rata Relevansi: 0.4667, Nilai Tertinggi Relevansi: 1.0000  
 Rata-Rata Senioritas: 0.7667, Nilai Tertinggi Senioritas: 0.8000



Nilai Korelasi untuk Spare part Admin: [-0.6000000000000001, -0.5, 0.9, -0.6000000000000001, -0.3999999999999999, -0.3000000000000004]  
 Persentase Relevance untuk Spare part Admin: [0.0, 0.0, 0.0, 0.4, 0.0, 0.2]  
 Persentase Senioritas untuk Spare part Admin: [0.4, 0.8, 0.6, 0.4, 0.8, 0.6]  
 Rata-Rata Korelasi: -0.2500, Nilai Tertinggi Korelasi: 0.9000  
 Rata-Rata Relevansi: 0.1000, Nilai Tertinggi Relevansi: 0.4000  
 Rata-Rata Senioritas: 0.6000, Nilai Tertinggi Senioritas: 0.8000



Nilai Korelasi untuk Teachers: [0.8, -0.5, -0.7, -0.7, -0.5, -0.5]  
 Persentase Relevance untuk Teachers: [0.8, 0.8, 1.0, 1.0, 0.8, 0.8]  
 Persentase Senioritas untuk Teachers: [1.0, 1.0, 0.4, 1.0, 1.0, 0.6]  
 Rata-Rata Korelasi: -0.3500, Nilai Tertinggi Korelasi: 0.8000  
 Rata-Rata Relevansi: 0.8667, Nilai Tertinggi Relevansi: 1.0000  
 Rata-Rata Senioritas: 0.8333, Nilai Tertinggi Senioritas: 1.0000



Nilai Korelasi untuk Unmanaged Merchant Engagement Senior Associate, BPO Field Sales: [0.0, 0.6, 0.7, 0.7, 0.7, 0.7]  
 Persentase Relevance untuk Unmanaged Merchant Engagement Senior Associate, BPO Field Sales: [0.4, 0.4, 0.4, 0.6, 0.8]  
 Persentase Senioritas untuk Unmanaged Merchant Engagement Senior Associate, BPO Field Sales: [0.2, 0.2, 0.4, 0.4, 0.2, 0.8]  
 Rata-Rata Korelasi: 0.5667, Nilai Tertinggi Korelasi: 0.7000  
 Rata-Rata Relevansi: 0.5000, Nilai Tertinggi Relevansi: 0.8000  
 Rata-Rata Senioritas: 0.3667, Nilai Tertinggi Senioritas: 0.8000

