

BAB V

ANALISA HASIL DAN PEMBAHASAN

Pada BAB V ini dibahas mengenai analisa hasil dan pembahasan dari skenario pengujian. Pada BAB V ini berisi bagaimana skenario pengujian, hasil pengujian, dan analisa hasil.

5.1. Skenario Pengujian

Skenario pengujian dibagi menjadi dua. Pengujian pertama membandingkan hasil pengujian algoritma *levenshtein distance* menggunakan bantuan *case folding*, *tokenizing*, *filtering*, dan *stemming* dengan membandingkan pengujian menggunakan algoritma *levenshtein distance* tanpa menggunakan bantuan *stemming* kemudian membandingkan nilai *similarity* yang didapat oleh sistem. Hal ini dilakukan untuk mengetahui seberapa besar efek dari *preprocessing* dalam melakukan penilaian jawaban ujian berbentuk esai. Kedua adalah pengujian untuk mengetahui pengaruh implementasi algoritma *levenshtein distance* yaitu dengan membandingkan penilaian manual (*human rater*) dengan penilaian otomatis. Pengujian ini bertujuan untuk mengetahui kerja algoritma *levenshtein distance* terhadap kemampuan sistem untuk menilai jawaban esai dibandingkan dengan penilaian manual (*human rater*). Penilaian akan didasarkan pada besarnya rata-rata selisih antara penilaian esai menggunakan algoritma *levenshtein distance* dengan penilaian manual (*human rater*). Pada skenario pengujian ini diambil 20 sampel mahasiswa dengan tiap mahasiswa 5 soal pengujian. Ujian sampel yang digunakan bersifat ujian esai dan sudah tersedia kunci jawaban dari masing-masing soal tersebut.

5.2 Hasil Pengujian

Pada pengujian ini, jumlah soal yang diujikan sebanyak 5 soal dengan materi yang diujikan adalah arsitektur organisasi komputer. Jumlah mahasiswa

yang diikutsertakan dalam skenario ini sebanyak 20 mahasiswa. Pengujian yang pertama membandingkan nilai dari sistem yang menggunakan proses *stemming* dan tanpa *stemming*. *Stemming* yang digunakan adalah stemming Arifin. Dengan pengujian tersebut, maka akan dapat dilihat perbedaan akurasi *similarity* dari perbandingan penilaian sistem dengan proses stemming dan tanpa stemming kemudian dilakukan juga pengujian penilaian. Pengujian ini ditujukan untuk mengamati perbedaan penilaian yang diimplementasikan di dalam sistem penilaian otomatis jawaban esai sehingga dapat diketahui perbedaan nilai sistem pada saat menggunakan algoritma *levenshtein distance* dan pengaruh *stemming*. Contoh pengujian perhitungan menggunakan data mahasiswa nomor 2. Berdasarkan contoh data latihan mempunyai perhitungan sebagai berikut:

Kunci1 : sebuah program yang mengamati dan mengatur serta mengendalikan atau memverifikasi operasi suatu sistem pengolahan data

Jawaban1 : urutan instruksi yang komputer dapat menafsirkan dan melaksanakan

- Nilai dosen: 20
- Nilai sistem dengan *stemming*: 5,161
- *Edit distance* / *Similarity*: 46 / 25,81%
- Nilai sistem tanpa *stemming*: 5,195
- *Edit distance* / *Similarity*: 57 / 25,97%

Kunci2 : proses yang sudah diatur dengan ketentuannya disebut sistem operasi agar tidak terjadi suatu tabrakan dengan proses lain

Jawaban2 : kunci untuk melakukan multiprogram dalam jangka panjang, jangka pendek, jangka menengah yang dilakukan di input output dengan melakukan penjadwalan

- Nilai dosen: 20
- Nilai sistem dengan *stemming*: 4.545
- *Edit distance* / *Similarity*: 68 / 22.73%
- Nilai sistem tanpa *stemming*: 4,909
- *Edit distance* / *Similarity*: 83 / 24,55%

Kunci3 : suatu proses yang dapat dialihkan sementara dari memori ke suatu tempat penyimpanan dan dipanggil kembali ke memori jika akan melanjutkan eksekusi

Jawaban3 : suatu proses yang dapat dialihkan sementara di memori ke suatu tempat penyimpanan dan dipanggil kembali ke memori jika akan melanjutkan eksekusi

- Nilai dosen: 20
- Nilai sistem dengan *stemming*: 20
- *Edit distance* / *Similarity*: 0 / 100%
- Nilai sistem tanpa *stemming*: 20
- *Edit distance* / *Similarity*: 0 / 100%

Kunci4 : bagian suatu program atau potongan program

Jawaban4 : panjang blok yang tetap memiliki alamat virtual yang di transfer

- Nilai dosen: 0
- Nilai sistem dengan *stemming*: 5
- *Edit distance* / *Similarity*: 30 / 25%
- Nilai sistem tanpa *stemming*: 4,444
- *Edit distance* / *Similarity*: 35 / 22,22%

Kunci5 : suatu bentuk bagian kerangka potongan yang tersedia di memori atau suatu area penyimpanan utama

Jawaban5 : suatu bentuk bagian kerangka potongan yang tersedia di memori / suatu area penyimpanan utama

- Nilai dosen: 20
- Nilai Sistem dengan *stemming*: 20
- *Edit distance* / *Similarity*: 0 / 100%
- Nilai sistem tanpa *stemming*: 20
- *Edit distance* / *Similarity*: 0 / 100%

Dari perbandingan jawaban diatas didapatkan nilai dari penilaian sistem dengan menjumlahkan semua hasil penilaian yang didapatkan. Sehingga hasilnya adalah:

Total nilai sistem dengan *stemming* : $5,161 + 4,545 + 20 + 5 + 20 = 54,706$

Total nilai sistem tanpa *stemming* : $5,195 + 4,909 + 20 + 4,444 + 20 = 54,548$

Nilai yang dihasilkan dari pengujian ini terbagi dari dua hasil, yaitu nilai yang dipengaruhi oleh *stemming* dan yang tidak dipengaruhi oleh *stemming*. Nilai hasil pengujian dengan *stemming* bernilai 54,706 dan yang tanpa *stemming* adalah 54,548. Kemudian dari hasil perhitungan penilaian sistem dengan *stemming* dan tanpa *stemming* dapat diperoleh prosentase kesalahan dari sistem.

Prosentase kesalahan sistem merupakan nilai maksimal jawaban dikurangi hasil bagi nilai sistem dengan nilai manual dikalikan 100%. Berikut adalah detail perhitungan manual prosentase kesalahan sistem :

$$\text{Prosentase kesalahan} = 100 - \left\{ \frac{NS}{NM} * 100 \right\}$$

dimana : NS = Nilai sistem

NM = Nilai manual

Prosentase kesalahan sistem dengan *stemming* :

$$= 100 - \left\{ \frac{54,706}{80} * 100 \right\}$$

$$= 31,62 \%$$

Prosentase kesalahan sistem tanpa *stemming* :

$$= 100 - \left\{ \frac{54,548}{80} * 100 \right\}$$

$$= 31,82 \%$$

Hasil penilaian dengan sistem akan dilakukan pembuktian dengan melakukan perhitungan secara manual. Hal ini untuk membuktikan apakah penilaian sistem melakukan perhitungan dengan baik atau belum. Berikut perhitungan manual:

Nomor 1

Kunci jawaban: sebuah program yang mengamati dan mengatur serta mengendalikan atau memverifikasi operasi suatu sistem pengolahan data

Jawaban: urutan instruksi yang komputer dapat menafsirkan dan melaksanakan

Hasil setelah dilakukan *preprocessing*

Kunci jawaban :

amataturdatamemverifikasimengendalikanolahoperasiprogramsisitem

Jawaban : instruksikomputerlaksanamenaafsirkanurut

$$\text{Similarity} = \left\{ 1 - \frac{46}{62} \right\} * 100$$

$$= 25,81 \%$$

$$\begin{aligned}\text{Nilai} &= \left\{1 - \frac{46}{62}\right\} * 20 \\ &= 5,161\end{aligned}$$

Nomor 2

Kunci jawaban: proses yang sudah diatur dengan ketentuannya disebut sistem operasi agar tidak terjadi suatu tabrakan dengan proses lain

Jawaban: kunci untuk melakukan multiprogram dalam jangka panjang, jangka pendek, jangka menengah yang dilakukan di input output dengan melakukan penjadwalan

Hasil setelah dilakukan *preprocessing*

Kunci jawaban: aturdisebuahjadioperasiprosesprosesistemtabraktentu

Jawaban:

inputjadwaljangkajangkajangkakuncilakulakulakumultiprogramoutputpanjangpendektenengahuntuk

$$\begin{aligned}\text{Similarity} &= \left\{1 - \frac{68}{88}\right\} * 100 \\ &= 22.73 \%\end{aligned}$$

$$\begin{aligned}\text{Nilai} &= \left\{1 - \frac{68}{88}\right\} * 20 \\ &= 4.545\end{aligned}$$

Nomor 3

Kunci jawaban: suatu proses yang dapat dialihkan sementara dari memori ke suatu tempat penyimpanan dan dipanggil kembali ke memori jika akan melanjutkan eksekusi

Jawaban: suatu proses yang dapat dialihkan sementara di memori ke suatu tempat penyimpanan dan dipanggil kembali ke memori jika akan melanjutkan eksekusi

Hasil setelah dilakukan *preprocessing*

Kunci jawaban:

aliheksekusikembalilanjutmemorimemori Panggil proses simpan tempat

Jawaban: aliheksekusikembalilanjutmemorimemori Panggil proses simpan tempat

$$\text{Similarity} = \left\{1 - \frac{0}{62}\right\} * 100$$

$$= 100 \%$$

$$\begin{aligned} \text{Nilai} &= \left\{ 1 - \frac{0}{62} \right\} * 20 \\ &= 20 \end{aligned}$$

Nomor 4

Kunci jawaban: bagian suatu program atau potongan program

Jawaban: panjang blok yang tetap memiliki alamat virtual yang di transfer

Hasil setelah dilakukan *preprocessing*

Kunci jawaban: bagipotongprogramprogram

Jawaban: alamatblokikipanjangtetaptransfervirtual

$$\begin{aligned} \text{Similarity} &= \left\{ 1 - \frac{30}{40} \right\} * 100 \\ &= 25 \%$$

$$\begin{aligned} \text{Nilai} &= \left\{ 1 - \frac{30}{40} \right\} * 20 \\ &= 5 \end{aligned}$$

Nomor 5

Kunci jawaban: suatu bentuk bagian kerangka potongan yang tersedia di memori atau suatu area penyimpanan utama

Jawaban: suatu bentuk bagian kerangka potongan yang tersedia di memori / suatu area penyimpanan utama

Hasil setelah dilakukan *preprocessing*

Kunci jawaban: areabagibentukkerangkamemoriopotongsediasimpanutama

Jawaban: areabagibentukkerangkamemoriopotongsediasimpanutama

$$\begin{aligned} \text{Similarity} &= \left\{ 1 - \frac{0}{50} \right\} * 100 \\ &= 100 \%$$

$$\begin{aligned} \text{Nilai} &= \left\{ 1 - \frac{0}{50} \right\} * 20 \\ &= 20 \end{aligned}$$

Hasil selengkapnya untuk pengujian semua mahasiswa ditunjukkan pada tabel 5.1

Tabel 5.1 Data Uji Tiap Mahasiswa

Peserta	Penilaian Dosen(a)	Penilaian Sistem dengan stemming(b)	Penilaian Sistem tanpa stemming(c)	Kesalahan (a) dan (b) (%)	Kesalahan (a) dan (c) (%)
Mahasiswa 1	80	36.564	37.99	54.3	52.51
Mahasiswa 2	80	54.706	54.548	31.62	31.82
Mahasiswa 3	60	35.968	37.696	40.05	37.17
Mahasiswa 4	60	43.543	45.999	27.43	23.34
Mahasiswa 5	60	39.116	33.696	34.81	43.84
Mahasiswa 6	100	84.222	84.643	15.78	15.36
Mahasiswa 7	60	65.789	67.429	9.65	12.38
Mahasiswa 8	80	26.201	26.859	67.25	66.43
Mahasiswa 9	80	40.369	41.277	49.54	48.4
Mahasiswa 10	80	30.403	30.521	62	61.85
Mahasiswa 11	80	51.511	57.058	35.61	28.68
Mahasiswa 12	100	36.148	40.898	63.85	59.1
Mahasiswa 13	80	52.92	58.279	33.85	27.15
Mahasiswa 14	40	35.319	53.403	11.7	33.51
Mahasiswa 15	40	50.306	40.034	25.77	0.09
Mahasiswa 16	60	55.163	55.617	8.06	7.31
Mahasiswa 17	80	37.42	34.233	53.23	57.21
Mahasiswa 18	80	43.972	39.119	45.04	51.1
Mahasiswa 19	80	41.296	42.435	48.38	46.96
Mahasiswa 20	60	28.707	31.911	52.16	46.82
Rata-rata				38.50	37.55

Sumber : [Rancangan]

Pengujian berikutnya dilakukan dengan menjumlahkan soal 1 sampai 5 dari setiap mahasiswa sehingga menghasilkan nilai seperti pada tabel 5.1. Prosentase kesalahan paling tinggi yang dihasilkan antara penilaian dosen /

penilaian manual dengan penilaian sistem menggunakan *stemming* adalah 67,25% dan prosentase kesalahan paling rendah adalah 8,06%. Sedangkan pada penilaian sistem tanpa *stemming* dengan penilaian manual prosentase kesalahan tertinggi adalah 66,43% dan paling rendah adalah 0,09%. Sehingga didapatkan selisih rata-rata dari kesalahan sistem dibandingkan dengan penilaian manual besar. Dengan hasil pengujian yang diperoleh bisa dikatakan bahwa sistem penilaian esai menggunakan algoritma *levenshtein distance* tanpa *stemming* memberikan penilaian yang lebih baik, walaupun selisihnya tidak jauh berbeda. Hal ini dikarenakan proses tanpa *stemming* tidak membaca perubahan imbuhan dalam kata sehingga berpengaruh pada hasil prosentase kemiripan jawaban dan juga pada nilai yang dihasilkan.

Hasil dari pengujian antara penilaian sistem menggunakan *stemming* dan tanpa *stemming* dapat dilihat pada lampiran. Terdapat selisih penilaian antara sistem penilaian yang menggunakan *stemming* dan tanpa *stemming* yaitu antara 0 – 3,938. Selisih terkecil dari kedua sistem antara penilaian yang menggunakan *stemming* dengan yang tanpa *stemming* didapat dari penilaian esai untuk beberapa mahasiswa yang jawaban dengan kunci jawaban mempunyai kemiripan 100%. Dari hasil percobaan terlihat tidak tampak perbedaan yang signifikan antara nilai kemiripan jawaban menggunakan *stemming* dibandingkan dengan tanpa *stemming*. Pada percobaan ini algoritma *levenshtein distance* dengan *stemming* memiliki kemiripan jawaban yang hampir sama dengan nilai algoritma *levenshtein distance* tanpa *stemming*.

5.3 Analisa Hasil

Berdasarkan hasil percobaan yang telah dilakukan terhadap seluruh data latih pada tabel 5.1 dan lampiran, perbandingan antara penilaian sistem menggunakan algoritma *levenshtein distance* dengan *stemming* dan tanpa *stemming* didapat hasil yang hampir sama sehingga mempunyai selisih minimal adalah 0 dan selisih maksimal adalah 3,938. Hal ini dikarenakan perubahan imbuhan tidak berpengaruh terlalu besar pada sistem ini.

Rata-rata prosentase kesalahan sistem yang didapat mencapai 38,50% dari hasil prosentase maka penilaian antara sistem dengan penilaian manual terjadi

perbedaan yang cukup signifikan. Hal ini bisa dikatakan kurang efektif jika dilihat besarnya selisih yang banyak terjadi pada penilaian esai. Selisih yang jauh didapatkan karena sistem masih belum bisa mendeteksi sinonim kata-kata yang berbeda tetapi memiliki makna yang sama. Jadi jika ada jawaban yang mempunyai arti sama tetapi didalam kunci jawaban tidak ada kemiripan katanya maka hal tersebut memiliki nilai yang rendah. Contoh pada seperti pada jawaban mahasiswa

Pertanyaan

apa yang dimaksud dengan scheduling?

Kunci jawaban

proses yang sudah diatur dengan ketentuannya disebuah sistem operasi agar tidak terjadi suatu tabrakan dengan proses lain

Jawaban

sebuah metode dari sistem operasi yang mengatur proses-proses yang akan berjalan dalam sistem operasi itu sendiri

Setelah dilakukan preprocessing

Kunci jawaban : atur disebuah jadi operasi proses sistem tabrakan tentu
jawaban : atur jalan metode operasi operasi proses proses sistem sistem

Perhitungan menggunakan algoritma levenshtein distance:

$$\begin{aligned} \text{- Nilai sistem} &= \left\{ 1 - \frac{24}{53} \right\} * 20 \\ &= 10,943 \end{aligned}$$

- Nilai manual: 20

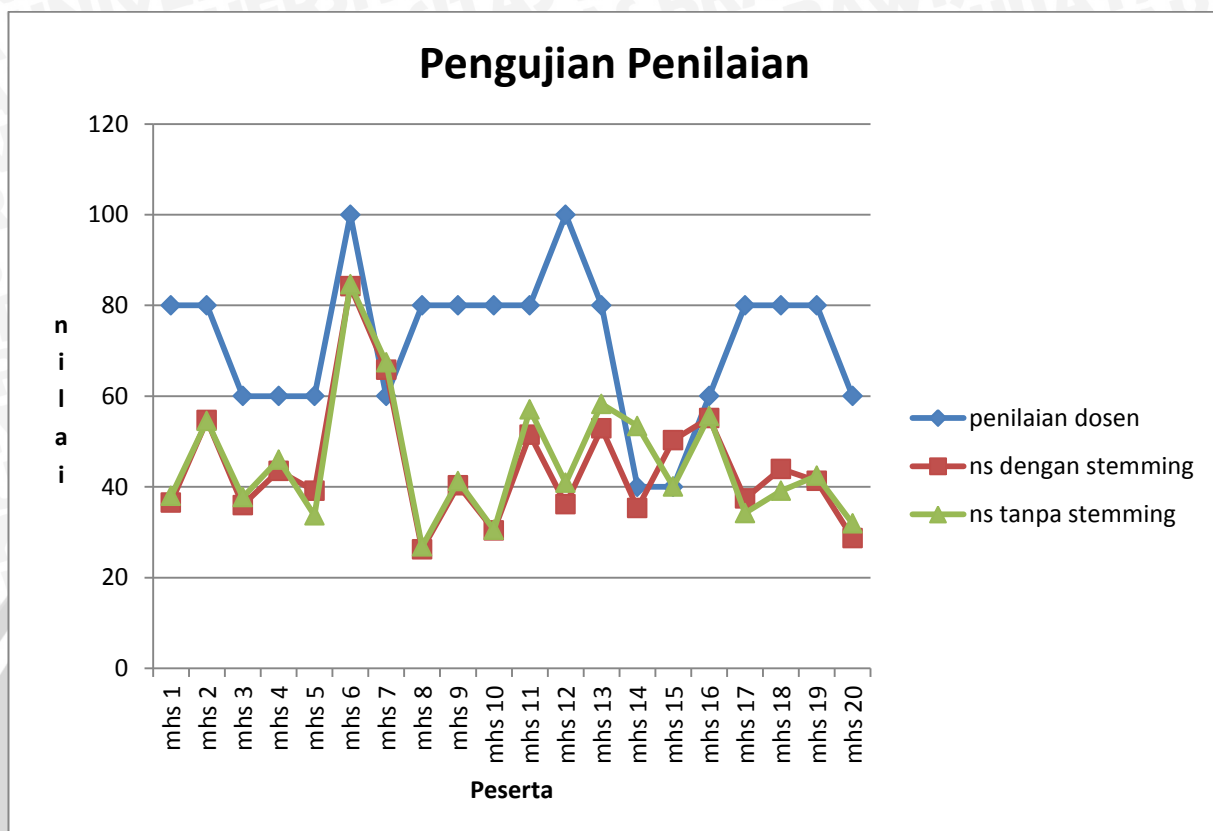
- *Edit distance* / *Similarity*: 24 / 54,72 %

Pada contoh penilaian manual memberikan nilai 20 karena jawaban dianggap benar karena bobot pada soal adalah 20 sedangkan nilai pada sistem 10.943 dengan tingkat kemiripan jawaban 54.72 %. Selisih yang didapatkan 9,057 antara penilaian sistem dengan penilaian manual. Dalam hal ini jawaban mahasiswa memiliki pengertian yang sama dengan kunci jawaban tetapi sistem

tidak dapat mendeteksi sinonim kata-kata yang berbeda tetapi memiliki makna yang sama.

Kemudian sistem masih belum bisa mendeteksi makna kalimat negatif dan positif dengan menggunakan antonim kata sehingga dapat lebih meningkatkan penilaian walaupun kemiripan pada jawaban berbeda. Karena konsep pada penilaian esai menggunakan algoritma *levenshtein distance* ini mencari kata *keyword* pada kunci jawaban dan jawaban mahasiswa kemudian dicocokkan. Nilai akan 100 jika kata yang ada pada kunci jawaban dan jawaban tersebut mempunyai kemiripan kata 100%. Pada tabel 5.1 hanya terdapat beberapa peserta yang nilainya mendekati sama antara penilaian sistem dengan penilaian dosen. Terdapat beberapa faktor-faktor yang mempengaruhi hasil penilaian yang didapatkan yaitu peserta ujian diberikan satu referensi yang dimana kunci jawaban nantinya merujuk pada jawaban peserta. Sehingga peserta dapat menjawab sesuai dengan referensi yang diberikan sebelumnya jadi jawaban nantinya tidak bersifat opini melainkan berupa *definitif* / sudah pasti.

Sistem akan mendapatkan akurasi yang baik jika sesuai dengan referensi yang sama sehingga jawaban peserta dan kunci jawaban memiliki kesamaan (*similarity*) maka nilai yang dihasilkan akan mendekati 100%. Grafik pengujian penilaian pada gambar 5.1 menyimpulkan penilaian sistem dengan menggunakan *stemming* hampir sama hasilnya dengan penilaian sistem tanpa menggunakan *stemming*. Sedangkan penilaian sistem dengan penilaian manual (*human raters*) terdapat perbedaan nilai walaupun beberapa masih ada nilai yang mendekati.



Gambar 5.1 Grafik Pengujian Penilaian

Sumber : [Rancangan]

Dari grafik pada gambar 5.1 dapat dianalisis hasilnya bahwa terdapat perbedaan antara penilaian manual dengan sistem yang cukup tinggi. Hal tersebut karena terdapat perbedaan standart penilaian antara manual dan sistem. Pada sistem jawaban yang bernilai tinggi adalah jawaban yang mengandung *string* hasil *preprocessing* kata kunci jawaban yang diberikan dosen. Serta memiliki kemiripan kunci jawaban dengan jawaban tiap kata yang hampir sama. Sedangkan pada penilaian manual/dosen, ada penilaian yang sudah dianggap benar meskipun jawaban tidak sama dengan kunci. Karena memiliki maksud yang sama dengan jawaban yang diberikan dosen. Pada penilaian manual jika ada jawaban mahasiswa yang kurang lengkap pada suatu kalimat bisa bernilai benar / bernilai tinggi. Hal ini karena penilaian manual tidak melihat jawaban secara keseluruhan. Pada penilaian manual jika terdapat kalimat jawaban yang benar pada jawaban mahasiswa, maka jawaban akan mendapatkan nilai yang tinggi. Sedangkan pada

sistem ini panjang pendeknya *string* pada jawaban berpengaruh terhadap penilaian yang diberikan. Panjang pendek pada kalimat berpengaruh pada penilaian semakin banyak perbedaan jumlah kata pada suatu kalimat antara jawaban dosen dengan mahasiswa, maka nilai yang dihasilkan semakin kecil. Hal lain yang menyebabkan perbedaan adalah kunci jawaban yang dimasukkan ke dalam sistem dapat memiliki persamaan pengertian dengan kata-kata lain yang mungkin saja menjadi jawaban mahasiswa. Misalkan persamaan/sinonim atau penulisan bahasa asing.

