

LAPORAN SKRIPSI

**KOMBINASI *POS TAGGING* DAN METODE *COSINE SIMILARITY*
UNTUK PENILAIAN JAWABAN *ESSAY* OTOMATIS**

(Studi Kasus : Mata Pelajaran IPA Biologi kelas IX SMPN 1 Kalianget)



Disusun Oleh :

MARETA KURNIA SARI

180411100090

Dosen Pembimbing 1 : Husni, S.Kom., M.T

Dosen Pembimbing 2 : Dr.Fika Hastarita Rachman, S.T., M.Eng.,

**PROGRAM STUDI TEKNIK INFORMATIKA
JURUSAN TEKNIK INFORMATIKA
FAKULTAS TEKNIK
UNIVERSITAS TRUNOJOYO MADURA
2022**

HALAMAN JUDUL

**KOMBINASI *POS TAGGING* DAN METODE *COSINE SIMILARITY*
UNTUK PENILAIAN JAWABAN *ESSAY* OTOMATIS**

(Studi Kasus : Mata Pelajaran IPA Biologi kelas IX SMPN 1 Kalianget)

SKRIPSI

Diajukan untuk Memenuhi Persyaratan Penyelesaian

Studi Strata Satu (S1) dan Memperoleh Gelar Sarjana Komputer (S.Kom)

di Universitas Trunojoyo Madura

MARETA KURNIA SARI

(18.04.1.1.1.00090)

PROGRAM STUDI TEKNIK INFORMATIKA

JURUSAN TEKNIK INFORMATIKA

FAKULTAS TEKNIK

UNIVERSITAS TRUNOJOYO MADURA

2021

HALAMAN PERNYATAAN ORISINALITAS

Saya yang bertanda tangan di bawah ini, menyatakan bahwa skripsi saya dengan judul:

“ KOMBINASI *POS TAGGING* DAN METODE *COSINE SIMILARITY* UNTUK PENILAIAN JAWABAN *ESSAY* OTOMATIS (Studi Kasus : Mata Pelajaran IPA Biologi kelas IX SMPN 1 Kalianget) ”

1. Adalah asli, bukan merupakan karya pihak lain serta belum pernah diajukan untuk mendapatkan gelar akademik Sarjana Komputer baik di Universitas Trunojoyo Madura maupun di Perguruan Tinggi yang lain di Indonesia.
2. Tidak terdapat karya atau pendapat pihak lain yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis telah diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Apabila dikemudian hari terbukti skripsi ini sebagian atau seluruhnya merupakan hasil plagiasi atau terdapat hal-hal yang tidak sesuai dengan pernyataan di atas, maka saya sanggup menerima sanksi akademis yang berlaku dengan segala akibat hukumnya sesuai peraturan Universitas Trunojoyo Madura dan atau peraturan perundang-undangan yang berlaku.

Bangkalan, Juni 2022

Yang Menyatakan,

Mareta Kurnia Sari

NIM. 180411100090

HALAMAN PENGESAHAN

**KOMBINASI *POS TAGGING* DAN METODE *COSINE SIMILARITY*
UNTUK PENILAIAN JAWABAN *ESSAY* OTOMATIS**

(Studi Kasus : Mata Pelajaran IPA Biologi kelas IX SMPN 1 Kalianget)

**Skripsi ini disusun untuk memenuhi salah satu syarat memperoleh gelar
Sarjana Komputer (S.Kom) di Universitas Trunojoyo Madura**

Oleh :

Nama : Mareta Kurnia Sari

NIM : 18.04.1.1.1.00090

Disetujui oleh Tim Penguji Skripsi :

Tanggal Sidang :

Husni, S.Kom., M.T

(Pembimbing I)

NIP. 19770722 200312 1 006

Dr.Fika Hastarita Rachman, S.T,M.Eng.,

(Pembimbing II)

NIP. 19830305 200604 2 002

Firdaus Solihin, S.Kom., M.Kom

(Penguji I)

NIP. 19760627 200801 1 008

Ika Oktavia Suzanti, S.Kom., M.Cs

(Penguji II)

NIP. 19881018 201504 2 004

Achmad Yasid, S.Kom., M.Kom

(Penguji III)

NIP. 19760627 200801 1 008

Bangkalan, Juni 2022

Mengetahui,

Ketua Jurusan Teknik Informatika

Dr. Yeni Kutsiyahningsih, S.Kom., M.Kom.

NIP. 19770921 200812 2 002

ABSTRAK

Ujian bertipe *essay* merupakan sebuah pilihan evaluasi guru untuk mengukur tingkat kemampuan berpikir siswa, karena dituntut untuk mengisi jawaban dengan kemampuan menguraikan kalimat siswa sendiri. Oleh karena itu banyak penelitian mengenai penilaian jawaban *essay* otomatis, tetapi penelitian sebelumnya tidak ada yang memperhatikan ambiguitas kata yang terdapat pada jawaban soal *essay*. Sistem penilaian jawaban *essay* otomatis yang akurat diperlukan untuk membantu penilaian hasil evaluasi atau ujian siswa, selain itu diperlukan keakuratan ambiguitas kata, karena dalam satu kata dapat bermakna banyak atau lebih dari satu. Sehingga diusulkan penilaian jawaban *essay* otomatis menggunakan *POS Tagging* dan metode *Cosine Similarity* yang dapat menyelesaikan masalah tersebut. Dari hasil penelitian yang dilakukan di SMPN 1 Kalianget pada siswa kelas IX-A yang berjumlah 19 siswa dengan 5 soal dan kunci jawaban mata pelajaran IPA pada bidang Biologi didapatkan hasil akurasi aplikasi sistem penilaian jawaban *essay* otomatis menggunakan *POS Tagging* dan *Cosine Similarity* yaitu 79,0%.

Kata Kunci : *essay, POS Tagging, Cosine Similarity.*

KATA PENGANTAR

Alhamdulillah dengan menyebut nama Allah Subhanuhu Wa Ta'ala (SWT) karena atas berkat dan rahmatnya penulis dapat menyelesaikan skripsi dengan judul “Kombinasi *POS Tagging* Dan Metode *Cosine Similarity* Untuk Penilaian Jawaban *Essay* Otomatis” sebagai persyaratan untuk menyelesaikan Program Sarjana Fakultas Teknik dan Jurusan Teknik Informatika. Dengan terselesaikannya skripsi ini, penulis juga tidak lupa bersyukur dan berterimakasih yang sebesar-besarnya kepada :

1. Allah SWT yang telah melimpahkan berkat, rahmat dan kesehatan sehingga penulis dapat merasakan kenikmatan yang diberikan sampai saat ini.
2. Kedua orang tua serta seluruh keluarga besar penulis yang selalu memberi dukungan, semangat dan doa yang tiada hentinya.
3. Husni, S.Kom., M.T selaku dosen pembimbing I dan Dr.Fika Hastarita Rachman, S.T., M.Eng., selaku dosen pembimbing II. Terimakasih sudah banyak meluangkan waktu dan tenaga untuk membimbing dan memberi arahan kepada penulis selama penelitian skripsi ini.
4. Bapak dan Ibu Dosen Universitas Trunojoyo Madura khususnya di Fakultas Teknik Program Studi S1 Teknik Informatika yang telah membagikan ilmu dan waktunya kepada penulis selama masa perkuliahan.
5. Partner skripsi terbaik Norita Febriyanti dan Nurfalah Alfurqon yang sudah membantu dan memberi semangat positif kepada penulis.
6. Sahabat-sahabat terbaik Lailatul Badria Dian Astuti Aprilia, Risky Sulistiyani, Nur Khofifah, dan Rike Ayu Arista yang sudah memberikan dukungan dan semangat.
7. Teman-teman seperjuangan mahasiswa Teknik Informatika Angkatan 2018 yang tidak dapat disebutkan secara langsung, terimakasih atas kenangan yang sudah diberikan.
8. Terimakasih kepada diri sendiri yang sudah berjuang sampai pada titik ini, sehingga dapat menyelesaikan skripsi ini.

Penulis sepenuhnya sadar bahwa pengerjaan skripsi ini tidak luput dari kekurangan dan kesalahan. Karena hal tersebut, penulis mohon maaf yang sebesar-besarnya atas segala kekurangan dan kesalahan yang ada. Berbagai kritik dan saran yang membangun sangat penulis harapkan kepada rekan-rekan pembaca.

Bangkalan, 17 mei 2022

Penulis

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PERNYATAAN ORISINALITAS.....	ii
HALAMAN PENGESAHAN.....	iii
ABSTRAK	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vii
DAFTAR GAMBAR	ix
DAFTAR TABEL.....	x
DAFTAR RUMUS	xi
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Perumusan Masalah.....	3
1.2.1 Permasalahan	3
1.2.2 Solusi Permasalahan	3
1.2.3 Pertanyaan Penelitian.....	3
1.3 Tujuan dan Manfaat Penelitian.....	4
1.3.1 Tujuan	4
1.3.2 Manfaat	4
1.4 Batasan Masalah.....	4
1.5 Sistematika Proposal	4
BAB II.....	6
TINJAUAN PUSTAKA.....	6
2.1 Tes <i>Essay</i>	6

2.2 Teks <i>Preprocessing</i>	7
2.3 <i>Part Of Speech Tagging (POS Tagging)</i>	8
2.3.1 <i>Rule Based</i>	9
2.4 Pembobotan <i>Term Frequency Inverse Document Frequency (TF IDF)</i>	10
2.5 <i>Cosine Similarity</i>	11
2.6 Kata Ambigu	11
2.7 Penelitian Terkait	12
BAB III	18
METODE USULAN	18
3.1 Lokasi dan Waktu Penelitian	18
3.2 Arsitektur Sistem	18
3.3 DataSet	27
3.4 Skenario Pengujian	29
BAB IV	32
HASIL DAN PEMBAHASAN	32
4.1 Proses Pengembangan Sistem	32
4.2 Lingkungan Uji Coba	37
4.3 Hasil Skenario	37
BAB V	40
PENUTUP	40
5.1 Kesimpulan	40
5.2 Saran	40
REFERENSI	41
LAMPIRAN	43

DAFTAR GAMBAR

Gambar 3. 1 Arsitektur Umum	18
Gambar 3. 2 <i>Flowchart</i> Sistem	19
Gambar 3. 3 <i>Flowchart POS Tagging</i> metode <i>Rule Based</i>	21
Gambar 3. 4 Contoh Proses Pelabelan <i>POS Tagging</i>	22
Gambar 4. 1 Nilai Siswa	37
Gambar 4. 2 Grafik Akurasi.....	38

DAFTAR TABEL

Tabel 2. 1 Penelitian terkait	12
Tabel 3. 1 Proses <i>Case Folding</i>	20
Tabel 3. 2 Hasil <i>POS Tagging</i>	23
Tabel 3. 3 Proses <i>Filtering</i>	23
Tabel 3. 4 Proses <i>Stemming</i>	24
Tabel 3. 5 <i>Term</i> Dokumen <i>Matriks</i>	24
Tabel 3. 6 Hasil Perhitungan <i>IDF</i>	24
Tabel 3. 7 Hasil Perhitungan <i>TF-IDF</i>	25
Tabel 3. 8 Hasil perhitungan <i>Cosine Similarity</i>	26
Tabel 3. 9 Skoring	26
Tabel 3. 10 Soal dan jawaban dari buku Intan Pariwara.....	27
Tabel 3. 11 Tagset <i>POS Tagging</i>	28
Tabel 3. 12 Skenario Pengujian	31
Tabel 4. 1 Spesifikasi Perangkat Keras.....	37
Tabel 4. 2 Hasil Skenario Pengujian	38

DAFTAR RUMUS

Rumus 2. 1 TF.....	10
Rumus 2. 2 IDF.....	10
Rumus 2. 3 TF-IDF.....	11
Rumus 2. 4 Cosine Similarity	11
Rumus 3. 1 <i>RMSE</i>	30
Rumus 3. 2 Akurasi.....	31

BAB I

PENDAHULUAN

1.1 Latar Belakang

SMP NEGERI 1 Kalianget (SMPN 1 Kalianget) merupakan salah satu sekolah negeri yang berakreditasi A di Kabupaten Sumenep, yang beralamatkan di Jl. Raya Kalianget Desa Kalimo'ok Kecamatan Kalianget dengan jumlah siswa mencapai 919 anak. Seturut dengan peran Pemerintah yang mendorong penerapan teknologi informasi dalam kegiatan belajar mengajar, maka SMPN 1 Kalianget telah memiliki fasilitas *E-Learning*. Sistem *E-Learning* menggunakan *Learning Management System (LMS) Google Classroom* yang digunakan saat ini yang memiliki fitur bagi guru untuk memberikan materi, memberikan tugas atau ujian *online*, pengumpulan tugas dan menampilkan nilai. Ujian *essay* secara *online* merupakan ujian yang menuntut siswa untuk menjawab ujian dengan menguraikan kalimat mereka sendiri karena tidak ada pilihan jawaban dan metode ujian yang digunakan adalah *online*[1].

Bentuk tugas atau ujian saat ini dapat berupa pertanyaan yang memiliki pilihan jawaban atau berupa *essay*. Pada tugas atau ujian yang berupa pertanyaan pilihan ganda, guru cukup mudah mendapatkan hasil nilai siswa, karena cek jawaban siswa dapat dilakukan sistem secara otomatis. Dalam ujian *essay* siswa dituntut memiliki pemahaman yang baik pada suatu ilmu serta digunakan dalam mengukur seberapa jauh pemahaman mendalam pada suatu ilmu[1]. Namun untuk pertanyaan *essay*, guru harus melakukan cek jawaban siswa satu persatu agar dapat menentukan nilai. Penilaian ujian secara *essay* dipengaruhi oleh *subyektifitas* penilaian guru yang menyebabkan penilaian tidak berimbang[2].

Pada tahun 2021 penelitian terdahulu melakukan penelitian dengan membuat aplikasi penilaian *essay* otomatis menggunakan metode *Cosine Similarity* untuk membantu guru agar tidak memakan waktu lama dalam penilaian jawaban siswa dan mengurangi *subyektifitas* penilaian, dengan diperoleh akurasi rata-rata 81%[1]. Memiliki kelemahan ketidak mampuan dalam kata-kata ambigu, ambigu merupakan kata-kata yang mempunyai lebih dari satu arti atau makna contohnya kata “bisa” dapat bermakna bisa melakukan dan bisa ular. Ambiguitas

kata menimbulkan ketidak akuratan pelabelan kata yang memicu terhadap kesalahan dalam memaknai arti kata dalam suatu kalimat[3].

Solusi untuk menganalisis ambiguitas kata di dalam dokumen dan mengurangi tingkat subyektifitas penilaian *essay* dapat dilakukan dengan menerapkan sistem yang dapat menganalisis kata ambigu sehingga dapat memberi label pada setiap kata dan membandingkan dokumen kunci jawaban guru dengan dokumen jawaban dari siswa. Sistem yang diusulkan terlebih dahulu melakukan pelabelan atau *tag* berdasarkan jenis kata seperti kata kerja, kata benda dan lain sebagainya. Pelabelan kata pada penelitian ini menggunakan *Part-Of-Speech Tagging (POS Tagging)* dengan metode *Rule Based*. Hasil pelabelan setiap kata pada jawaban selanjutnya melakukan teks *preprocessing* untuk mendapatkan kata dari kalimat jawaban. Selanjutnya melakukan perhitungan *Term Frequency Inverse Document Frequency (TF-IDF)* yang selanjutnya akan dihitung dengan metode *Cosine Similarity* untuk mendapatkan tingkat kemiripan antara jawaban siswa dengan kunci jawaban. Metode *Cosine Similarity* dapat dijadikan pedoman dalam penelitian karena memiliki hasil kemiripan yang tepat[1].

Dari beberapa penelitian penilaian *essay* otomatis sebelumnya, belum ada yang memperhatikan ambiguitas kata yang terdapat pada jawaban soal *essay*, karena dalam interaksi manusia dengan komputer terdapat kajian mengenai *Natural Language Processing (NLP)*. Merupakan peranan *NLP* dalam mengurangi permasalahan ambiguitas kata yang terdapat pada teks berbahasa Indonesia[3]. Oleh sebab itu diadakan penelitian ini untuk mengetahui seberapa penting pengaruh ambiguitas kata pada penilaian *essay* otomatis, menggunakan *POS Tagging* dengan metode *Rule Based* untuk mendapatkan aturan ambiguitas kelas kata dan menggunakan metode *Cosine Similarity* untuk melakukan penilaian *essay* secara otomatis untuk mengurangi subyektifitas penilaian guru.

Penelitian terdahulu pada tahun 2018 melakukan penelitian penentuan kelas kata menggunakan *POS Tagging*, yang menggunakan korpus Bahasa Indonesia dan menggunakan metode *Rule Based*. Penelitian tersebut sudah berhasil menyusun algoritma penentuan ambiguitas kelas kata *POS Tagging* Bahasa Indonesia. Dengan hasil pengujian yaitu algoritma tersebut telah berhasil mendapatkan label 92 kata dari 100 kata ambigu dengan benar, dengan

perbandingan 2 kata tidak tepat dan 6 kata tidak dapat ditangani oleh algoritma[3]. Faktor-faktor yang mempengaruhi ketepatan algoritma adalah kelengkapan aturan, pelabelan kelas kata dan korpus yang digunakan.

Hasil penelitian ini diharapkan dapat membantu guru SMPN 1 Kalianget pada saat melakukan penilaian jawaban tugas atau ujian yang bersifat *essay*. Dengan sistem ini, Guru tidak perlu lagi memeriksa jawaban siswa secara satu persatu dan hasil penilaian dapat bersifat obyektif.

1.2 Perumusan Masalah

1.2.1 Permasalahan

Sistem Penilaian guru dalam soal berbentuk *essay* harus melakukan cek jawaban siswa satu persatu untuk menentukan penilaian dan dapat menyebabkan penilaian tidak berimbang. Oleh karena itu banyak penelitian mengenai penilaian jawaban *essay* otomatis, tetapi penelitian sebelumnya tidak ada yang memperhatikan ambiguitas kata yang terdapat pada jawaban soal *essay*. Sistem penilaian jawaban *essay* otomatis yang akurat diperlukan untuk membantu penilaian hasil evaluasi atau ujian siswa, selain itu diperlukan keakuratan ambiguitas kata, karena dalam satu kata dapat bermakna banyak atau lebih dari satu. Sehingga diusulkan penilaian jawaban *essay* otomatis menggunakan *POS Tagging* dan metode *Cosine Similarity* yang dapat menyelesaikan masalah tersebut.

1.2.2 Solusi Permasalahan

Solusi yang digunakan untuk mengurangi tingkat subyektifitas penilaian *essay*, merancang dan membuat penilaian *essay* otomatis menggunakan *POS Tagging* dengan metode *Rule Based* untuk mendapatkan aturan ambiguitas kelas kata dan menggunakan metode *Cosine Similarity* untuk melakukan perhitungan penilaian *essay* secara otomatis.

1.2.3 Pertanyaan Penelitian

Berapa nilai akurasi sistem penilaian *essay* otomatis jawaban siswa pada kombinasi *POS Tagging* dan metode *Cosine Similarity*?

1.3 Tujuan dan Manfaat Penelitian

1.3.1 Tujuan

Tujuan penelitian ini mengetahui seberapa akurat penggunaan kombinasi *POS Tagging* dan metode *Cosine Similarity* pada sistem penilaian jawaban *essay* otomatis.

1.3.2 Manfaat

Adapun manfaat dari kegiatan penelitian ini adalah sebagai berikut:

1. Guru tidak perlu lagi melakukan penilaian jawaban soal *essay* secara manual.
2. Hasil penilaian bersifat obyektif, karena dilakukan oleh sistem.

1.4 Batasan Masalah

Adapun Batasan-batasan masalah dari kegiatan penelitian ini adalah sebagai berikut:

1. Menggunakan 5 soal dan kunci jawaban mata pelajaran IPA (Ilmu Pengetahuan Alam) pada bidang Biologi dari bank soal kurikulum 2013 dan buku IPA kelas IX PT Intan Pariwara, dengan jawaban siswa kelas IX-A yang berjumlah 19 orang siswa SMPN 1 Kalianget.
2. Menggunakan data korpus Bahasa Indonesia dari Fakultas Ilmu Komputer Universitas Indonesia.
3. Setiap soal mempunyai bobot penilaian yang sama.
4. Menggunakan *library Stemming* dan *Stopword* dari Sastrawi, dan *Stemming* berbasis Nazief Adriani.

1.5 Sistematika Proposal

Sub bab ini berisikan urutan dari struktur penulisan laporan penelitian.

BAB I PENDAHULUAN

Bab ini berisi latar belakang, permasalahan penelitian, tujuan dan manfaat penelitian dan sistematika penulisan.

BAB II TINJAUAN PUSTAKA

Pada bab ini dijabarkan tentang jurnal penelitian yang terkait dan studi literatur yang menunjang dan mendukung terhadap permasalahan yang sedang dikaji.

BAB III METODE USULAN

Bab ini berisikan metode penelitian yang digunakan. Metode penelitian berisikan analisis kebutuhan, perancangan penelitian, teknik analisi.

BAB IV HASIL DAN PEMBAHASAN

Bab ini berisi tentang hasil implementasi sistem dengan menggunakan kombinasi *POS Tagging* dan metode *Cosine Similarity* dan hasil implementasi terdiri dari Proses pengembangan sistem, lingkungan uji coba dan hasil skenario.

BAB V PENUTUP

Bab ini berisi tentang kesimpulan dari penelitian yang disesuaikan dengan pertanyaan penelitian dan saran terkait dengan pengembangan aplikasi selanjutnya.

REFERENSI

Bab ini berisi sumber pustaka yang digunakan pada BAB I, BAB II, BAB III, BAB IV dan BAB V.

BAB II

TINJAUAN PUSTAKA

2.1 Tes *Essay*

Tes *essay* atau tes bentuk uraian adalah tes berupa pertanyannya mengharuskan jawaban berisi penjelasan berdasarkan pendapat dari peserta tes[1]. Dalam tes *essay* dituntut untuk dapat mengingat dan mengulas kembali, dan berkreativitas yang tinggi dalam menjawab soal. Tes *essay* atau tes uraian memiliki karakteristik sebagai berikut[4] :

- a. Bentuk tes sebuah pertanyaan atau perintah yang jawabannya berupa uraian atau sebuah penjelasan kalimat yang biasanya cukup panjang.
- b. Bentuk pertanyaan yang menuntut untuk memberikan penjelasan, komentar, tafsir, perbandingan, membedakan dan sebagainya
- c. Butir soal yang diberikan jumlahnya terbatas, yaitu antara lima sampai sepuluh butir soal.
- d. Butir soal pada tes *essay* biasanya diawali kata : “Jelaskan ...”, “Terangkan ...”, “Mengapa ...”, atau kata-kata serupa.

Tes *essay* merupakan pilihan evaluasi guru untuk mengukur tingkat kemampuan berpikir siswa, karena siswa dituntut untuk mengisi jawaban dengan kemampuan menguraikan, diskusi, membandingkan, memberi alasan, dan bentuk lain sesuai dari pertanyaan dengan kalimat mereka sendiri. Tes *essay* atau tes uraian memiliki kelebihan sebagai berikut :

- a. Mudah dalam mempersiapkan dan menyusunnya.
- b. Tidak ada kesempatan untuk siswa dalam berspekulasi menjawab soal.
- c. Mendorong siswa untuk berani menjawab soal dengan menyusun kalimat sendiri dengan bagus.
- d. Memberikan kesempatan untuk siswa dalam mengutarakan kalimat dengan Bahasa sendiri.
- e. Mengetahui pengetahuan siswa terhadap sejauh mana tingkat pemahaman materi, dan sebagainya.

Sedangkan tes *essay* atau tes uraian memiliki kekurangan sebagai berikut :

- a. Sampel tes *essay* yang terbatas sehingga tes *essay* tidak dapat menguji semua bahan ajar yang sudah diberikan, tidak untuk tes objektif yang dapat menguji banyak pertanyaan dengan pilihan jawaban.
- b. Sifatnya tes *essay* yang subjektif, dalam pertanyaan, dalam membuat pertanyaan, dan dalam cara memeriksanya.
- c. Dalam mengkoreksi tes *essay* diperlukan waktu yang lama, sehingga tidak praktis dan dapat menyebabkan penilaian menjadi subyektif.
- d. Pengoreksiannya dapat lebih sulit lagi karena dapat lebih banyak mempertimbangkan penilaian individual.
- e. Memakan waktu yang sangat lama untuk melakukan penilaian jawaban siswa dan tidak dapat diwakilkan.

2.2 Teks *Preprocessing*

Teks Preprocessing merupakan pengelompokan setiap kata yang sama pada dokumen satu dengan dokumen lainnya[1]. Berikut ini tahap *teks preprocessing* pada penelitian ini.

1) *Case Folding*

Case folding merupakan tahapan yang bertujuan mengubah semua huruf yang ada di dalam dokumen menjadi huruf kecil. Dan yang dapat diterima hanya huruf “a” sampai dengan “z”. karakter lain selain huruf akan dihilangkan seperti menghapus angka, menghapus tanda baca, dan menghapus *whitepace* atau karakter kosong.

2) *Tokenizing*

Tokenizing merupakan tahapan yang bertujuan sebagai proses memisahkan teks menjadi sebuah potongan-potongan atau dapat disebut token yang kemudian untuk dianalisa.

3) *Filtering*

Filtering dapat disebut juga *Stopword Removal* merupakan tahapan yang bertujuan untuk menghapus kata-kata dengan informasi rendah dan fokus pada kata-kata yang penting. Dengan penggunaan algoritma *stoplist* dalam pengambilan kata-kata yang penting dan *wordlist* untuk menghapus atau

membuang kata-kata yang cenderung kurang penting. Berikut contoh *stopword* Bahasa Indonesia adalah “yang”, “dan”, “dari”, dan lain-lain.

4) *Stemming*

Stemming merupakan tahapan yang bertujuan untuk menghilangkan infleksi pada kata menjadi bentuk kata dasarnya. Misal kata mendengarkan, didengar, dengarkan akan menjadi kata dasar yaitu “dengar”. *Stemming* umumnya lebih banyak digunakan pada teks berbahasa Inggris untuk *stemming* Bahasa Indonesia yang imbuhanannya sangat kompleks sehingga akan sedikit lebih sulit untuk diolah.

2.3 *Part Of Speech Tagging (POS Tagging)*

Part-Of-Speech Tagging atau dapat disebut *POS Tagging* adalah proses memberi label kategori *morphosyntactic* (contohnya *noun-singular*, *verb-past*, *adjective*, *pronoun-personal*, dan sebagainya) pada setiap kata di dalam teks[5]. Dalam *POS Tagging* sistem menerima teks yang sudah disegmentasi kalimatnya sebagai masukan. Pemanfaatan *Tagging* juga dalam aplikasi *Neuro Linguistic Programming (NLP)* seperti *Question and Answering (Q&A)*, *Teks Summarization*, *Machine Translation*, dan lainnya. Dalam proses *POS Tagging* diperlukan sebuah kamus atau korpus penggunaannya dalam penentuan kelas kata.

Berikut merupakan kelas kata Bahasa Indonesia sebagai berikut[6] :

a. Kata benda (*Noun*)

Kata benda yaitu kata atau gagasan kata sebagai pernyataan tentang sesuatu yaitu nama seseorang, nama tempat, binatang, sifat, ide dan perbuatan.

b. Kata kerja (*Verb*)

Kata kerja ialah kata atau gagasan kata yang menerangkan atau menggambarkan sebuah kejadian, tingkah laku, perbuatan, peristiwa dan keadaan.

c. Kata sifat (*Adjective*)

Kata sifat ialah kata atau gagasan kata yang digunakan untuk menerangkan kata benda, dengan mengubah kata benda dengan menjelaskannya.

d. Kata ganti orang (*Personal Pronoun*)

Kata ganti ialah kata atau gagasan kata yang digunakan untuk penggantian nama, seperti *first person* yaitu kata ganti orang pertama, *second person* yaitu kata ganti orang kedua, dan *third person* yaitu kata ganti orang ketiga.

e. Kata keterangan (*Adverb*)

Kata keterangan ialah kata atau gagasan kata sebagai pembatas atau pemberi informasi lebih banyak tentang kata kerja.

f. Kata bilangan (*Numeral*)

Kata bilangan ialah kata atau gagasan kata yang menunjukkan suatu bilangan. Kata bilangan dapat mendampingi kata bilangan yang lain.

g. Kata penghubung (*Conjunction word*)

Kata penghubung ialah kata atau gagasan kata yang memperluas satuan kata dan sebagai penghubung dengan beberapa satuan kata.

h. Kata sandang (*determiner*)

Kata sandang ialah kata atau gagasan kata yang menjadi pelengkap dan penjelas kata lain, biasanya kata sandang berada sebelum kata benda.

i. Kata seru atau Interjeksi (*Interjection*)

Kata seru ialah kata atau gagasan kata yang menunjukkan ungkapan rasa hati atau perasaan seseorang. Misalnya seperti kagum, sedih, heran, jijik dan sebagainya.

j. Kata depan atau Preposisi (*Preposition*)

Kata depan ialah kata atau gagasan kata yang posisi katanya ada di depan sebelum kata benda, kata kerja, dan kata keterangan lainnya.

Dalam *POS Tagging* dapat menggunakan beberapa metode yaitu *Rule Based*, *Hidden Markov Model (HMM)*, dan Algoritma Viterbi. Maka dari itu, penelitian ini menggunakan metode *Rule Based* dalam *POS Tagging* Bahasa Indonesia.

2.3.1 Rule Based

Metode *Rule Based* adalah metode penggunaannya memakai aturan bahasa (*grammar*) dalam mendapat sebuah aturan kelas kata pada suatu kata yang ada pada kalimat[7]. Arsitektur metode *Rule Based* ada 2 yaitu merupakan metode

Rule Based yang penggunaannya dengan kamus sebagai penanda kata dengan kelas kata (*leksikon*) dan metode *Rule Based* menggunakan *disambiguation rule* dengan manual untuk diproses selanjutnya dalam setiap kata menjadikan satu kelas kata saja. Kemudian dari kelas kata pertama sampai dengan kelas kata terakhir nantinya akan ada pencocokan *rule(aixan)* yang terdapat pada kamus aturan. Maka dalam susunan *rule* pada kalimat terdapat kamus aturan, jadi sistem tersebut akan menampilkan kata beserta kelas kata sebagai output. Jika dalam perbedaan kelas kata ditemukan dengan kelas kata dalam kamus, kemudian sistem langsung memberi peringatan atau tanda pada setiap kata tersebut dan pada kelas kata yang lebih benar dari kelas kata yang didapat maka akan ditampilkan oleh sistem.

2.4 Pembobotan *Term Frequency Inverse Document Frequency (TF IDF)*

Metode pembobotan *term (term wighting)* adalah dengan menggunakan frekuensi kemunculan *term* (kata) / *term frequency* atau *TF* yang bersangkutan pada suatu dokumen yang merupakan metode pembobotan paling sederhana. *TF* merupakan sebuah kuatitas *term* yang sering muncul pada suatu dokumen[1]. Proses untuk menghitung jumlah dari kemunculan (frekuensi) *term* t_i dalam setiap dokumen d_j [1].

$$W_{TF}(t_i, d_j) = f(t_i, d_j)$$

Rumus 2. 1 TF

Dimana:

$W_{TF}(t_i, d_j)$ = nilai *TF term* ke i pada dokumen ke j

$f(t_i, d_j)$ = jumlah kemunculan dari *term* ke i pada dokumen ke j

Inverse Document Frequency (IDF) adalah sebuah untuk mengetahui seberapa besar pengaruh *term* di dalam suatu dokumen terhadap dokumen lainnya. Dalam suatu dokumen dapat mengandung *term* yang sangat bernilai yang jarang sekali ada atau muncul[1].

$$W_{IDF} = 1 + \log \frac{D}{d(t_i)}$$

Rumus 2. 2 IDF

Dimana:

$W_{IDF}(t_i, d_j)$ = nilai *IDF term* ke i pada dokumen ke j

d_{t_i} = jumlah dokumen yang mengandung *term* ke i

D = Jumlah dokumen

Rumus untuk menyatakan bobot (w) dari dokumen yang diproses terhadap dokumen kunci adalah:

$$W_{tf-idf}(t_i, d_j) = W_{TF}(t_i, d_j) \times W_{IDF}(t_i, d_j) \quad \textbf{Rumus 2. 3 TF-IDF}$$

Dimana:

$W_{tf-idf}(t_i, d_j)$ = nilai *TF-IDF term* ke i pada dokumen ke j

$W_{IDF}(t_i, d_j)$ = nilai *IDF term* ke i pada dokumen ke j

$W_{TF}(t_i, d_j)$ = nilai *TF term* ke i pada dokumen ke j

2.5 Cosine Similarity

Metode *Cosine Similarity* adalah metode yang penggunaannya dalam menghitung tingkat kesamaan antara satu objek dengan objek yang lain[2]. Dalam penggunaan *Cosine Similarity* tujuannya ialah untuk membandingkan tingkat kecocokan antara dua objek, karena nilai *cosinus* 0° adalah 1 dan kurang dari 1 (<1) untuk nilai sudut yang lain. Maka suatu nilai kemiripan antar dua objek dapat dikatakan mirip ketika nilai dari *Cosine* adalah 1.

Cosine Similarity juga dapat diterapkan dalam penentuan nilai kemiripan pada dua dokumen teks. Dengan menggunakan parameter jumlah kata-kata pada dua dokumen teks yang membandingkan (misal D1 "Dokumen 1" dan D2 "Dokumen 2"). Berikut merupakan rumus umum *Cosine Similarity*[1].

$$CosSim(d_i, d_j) = \frac{q_i \cdot d_i}{|q_i| \cdot |d_i|} = \frac{\sum_{j=1}^t (q_{ij} \cdot d_{ij})}{\sqrt{\sum_{j=1}^t (q_{ij})^2 \cdot \sum_{j=1}^t (d_{ij})^2}} \quad \textbf{Rumus 2. 4 Cosine Similarity}$$

Dimana:

q_{ij} = bobot istilah j pada dokumen i = *TF-IDF*

d_{ij} = bobot istilah j pada dokumen i = *TF-IDF*

2.6 Kata Ambigu

Kata ambigu adalah kata yang mempunyai makna lebih dari satu[3]. Sehingga kata tersebut menimbulkan keraguan, ketidak jelasan dan sebagainya. Berikut terdapat 3 jenis ambigu[8], yaitu :

a. *Abiguitas Fonetik*

Merupakan jenis keambiguan akibat persamaan bunyi dan ucap.

b. *Ambiguitas Gramatikal*

Merupakan jenis keambiguan akibat proses terbentuknya tata bahasanya, akan tetapi kata ambiguitas *Gramatikal* ini akan hilang jika masuk dalam konteks kalimat.

c. *Ambiguitas Leksikal*

Merupakan jenis keambiguan akibat faktor kata itu sendiri. Dan pada penelitian ini menggunakan jenis ambiguitas *Leksikal*.

Dan dalam kata ambigu juga terdapat faktor penyebab ambigu[8], yaitu :

a. *Faktor Morfologi*

Faktor ambigu terjadi karena pembentukan kata itu sendiri.

b. *Faktor Sintaktis*

Faktor ambigu terjadi karena pada kalimat terdapat susunan kata yang kurang jelas.

c. *Faktor Struktural*

Faktor ambigu terjadi karena struktur dari kalimat itu sendiri. Dan pada penelitian ini menggunakan faktor *Struktural*.

2.7 Penelitian Terkait

Pada sub bab ini berisikan penelitian sebelumnya yang menjadi referensi kegiatan penelitian ini.

Tabel 2. 1 Penelitian terkait

Peneliti, Tahun	Permasalahan	Metode Yang Digunakan	Hasil Penelitian
Rahimi Fitri, Arifin Noor Asyikin, 2015[2]	Dalam penilaian jawaban <i>essay</i> memerlukan waktu yang lama atau menguras	<i>Cosine</i> <i>Similarity</i>	Sistem penilaian ujian <i>essay</i> secara otomatis dengan menerapkan metode <i>Cosine</i> <i>Similarity</i> . Berdasarkan

	waktu. Semakin banyak jumlah soal ujian dengan bertambah juga banyaknya siswa yang mengikuti ujian. Maka semakin banyak yang akan dinilai. Hal tersebut membuat menurunnya penilaian menjadi tidak objektif lagi.		hasil penelitian, metode <i>Cosine Similarity</i> telah berjalan dengan baik dengan tingkat kesesuaian nilai sistem dengan nilai yang diberikan oleh pengajar adalah rata-rata 89,48%
Ahmad Subhan Yasid, Agung Fatwanto, 2018[3].	Bahasa Indonesia merupakan Bahasa dengan urutan peringkat ke Sembilan dan memiliki permasalahan dalam penerapan mesin pembelajarannya, pada ambiguitas kata yang menimbulkan ketidakakuratan dalam penentuan label kata yang nantinya terdapat kesalahan dalam	<i>POS Tagging Rule Based</i>	Berhasil menyusun algoritma penentuan kelas kata ambigu pada <i>POS Tagging</i> Bahasa Indonesia dan dapat diimplementasikan dengan penerapan 71 aturan sintaktis. Yang berhasil memberi label 92 kata dari 100 kata ambigu dengan perbandingan 2 kata tidak benar dan 6 kta tidak ditangani oleh sistem. Kelengkapan aturan, label kelas kata dan korpus menjadi faktor pendukung,

	memaknai arti dari kalimat.		dengan akurasi (92%)
<i>KK Purnamasari, I S Suwardi, 2018[9].</i>	Dalam Bahasa Indonesia terdapat masalah yang ditemui yaitu kurangnya korpus beranotasi Bahasa Indonesia. Penelitian ini menggunakan korpus dengan KBBI (Kamus Besar Bahasa Indonesia) dengan beberapa aturan.	<i>POS Tagging Rule Based</i>	Sistem dibagi menjadi 4 Penelitian ini menghasilkan akurasi yang cukup baik yaitu (87,4%), tetapi deteksi POS Tagging belum dapat menyelesaikan kata-kata ambigu. Oleh karena itu pengembangan penelitian ini dapat menambah pola <i>morfologi</i> selang kata dan menambah kata ambiguitas.
Firmansyah Fataruba 2018[10].	Dalam mengoreksi jawaban dari siswa, guru membutuhkan waktu yang cukup lama sehingga bisa memengaruhi kualitas penilaian terhadap nilai siswa. Dalam penelitian ini menggunakan data ujian <i>essay</i> Biologi.	<i>Cosine Similarity</i>	Dari hasil pengujian yang dilakukan bahwa metode <i>Cosine Similarity</i> sebagai pengekan kemiripan jawaban siswa SMAN 11 Ambon maka dapat berjalan 100% pada dua <i>browser</i> dan dengan nilai akurasi mencapai 80% menggunakan <i>Confision Matrix</i> .
Nurul	Pendidikan di era	<i>Cosine</i>	Penelitian ini sudah

Hidayat, Lasmedi Afuan, 2019[7]	sekarang dilakukan dengan metode jarak jauh atau dapat disebut PJJ APTIKOM. Oleh sebab itu untuk penilain ujian online <i>essay</i> dapat objektif, dengan dikembangkan sistem ujian online yang berguna untuk penilaian jawaban ujian <i>essay</i> secara otomatis.	<i>Similarity</i>	berhasil mengembangkan sistem ujian <i>essay</i> otomatis pada PJJ APTIKOM. Dan diharapkan dapat membantu pendidik dalam penilaian ujian <i>essay</i> otomatis. Hasil penilaian jawaban bersifat <i>realtime</i> .
Mia Kamayani, 2019[11].	Mangidentifikasi <i>POS Tagging</i> Indonesia yang cukup sulit dibandingkan karena pada setiap penelitian menggunakan data set dan tagset yang berbeda. Tujuan penelitian ini untuk mengetahui dataset dan tagset yang layak digunakan sebagai	<i>Rule Based, Probabilistik, HMM dan Neural Network.</i>	Penelitian ini menghasilkan akurasi metode <i>NN 1-layer</i> (80%), <i>rulebase</i> (85,77%), <i>NN 2-layer</i> (94%), <i>CRF</i> (96,22%), <i>biLSTM+CRF</i> (97,47%).

	standar untuk melakukan <i>POS Tagging</i> Bahasa Indonesia.		
Eka Iarasati Amalia, Agelita Justien Jumadi, Irsyad Arif Mashudi, Dimas Wahyu Wibowo, 2021[1]	Dalam melakukan ujian <i>essay</i> online penilaian jawaban ujian yang memerlukan waktu yang cukup lama jika dikerjakan secara manual. Agar dalam proses penilaian tidak menguras waktu yang banyak maka di adakan penelitian ini untuk membuat aplikasi penilaian jawaban ujian online <i>essay</i> otomatis	<i>Cosine Similarity</i>	Berdasarkan hasil pengujian yang sudah dilakukan maka ditarik kesimpulan bahwa metode <i>Cosine Similarity</i> dapat dijadikan pedoman dalam penelitian karena memiliki hasil kemiripan yang tepat. Dengan rata-rata nilai akurasi 81%.
Nurlaily Vendyansyah, Yosep Agus Pranoto, 2021[12].	Banyaknya penilaian jawaban <i>essay</i> yang harus dinilai oleh guru, dapat menyebabkan menurunnya kualitas penilaian	<i>Cosine Similarity</i>	Pengujian penelitian ini dilakukan menggunakan 3 buah data uji, telah disimpulkan dengan selisih error, hasil aplikasi dengan hasil perhitungan manual yaitu sebesar 0,27 %.

	<p>dari guru dan penilaian yang cenderung subyektif dan juga dari segi waktu yang relative lama dalam penilaian. Penelitian ini menggunakan data dalam bidang ilmu Biologi yaitu Kingdom Plantae.</p>		<p>Aplikasi penelitian ini dapat dijalankan dengan baik di dua browser yang di ujikan yaitu <i>Google Chrome</i> dan <i>Mozilla Firefox</i>.</p>
--	---	--	--

Pada penelitian sebelumnya penilaian jawaban *essay* otomatis menggunakan metode *Cosine Similarity* sebagai penilaian kemiripan jawaban, dan *POS Tagging* dengan metode *Rule Based* untuk mendapatkan aturan ambiguitas kata dan juga melakukan pelabelan kata, karena jika urutan posisi *SPOK* (*Subjek, Predikat, Objek, Keterangan*) berubah maka makna kata juga dapat ikut berubah.

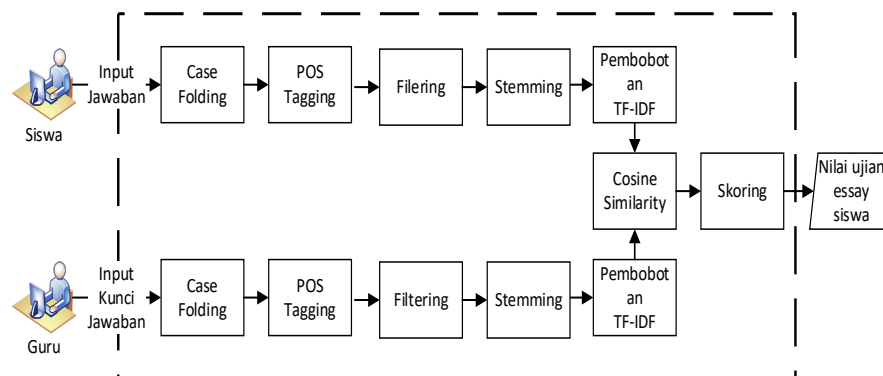
BAB III

METODE USULAN

3.1 Lokasi dan Waktu Penelitian

Lokasi penelitian ini bertempat di SMPN 1 Kalianget yang merupakan sekolah yang sudah berakreditasi A yang beralamatkan di JL. Raya Kalianget Desa Kalimo'ok Kecamatan Kalianget dengan jumlah siswa mencapai 919 anak. Penelitian ini dilakukan mulai dari September sampai dengan Oktober 2021.

3.2 Arsitektur Sistem



Gambar 3. 1 Arsitektur Umum

Berdasarkan rancangan arsitektur pada gambar 3.1 terdapat beberapa tahapan yang harus dilakukan dalam sistem :

a. Input

Data yang diinputkan ke dalam sistem terbagi menjadi dua bagian yaitu data yang dimasukkan siswa berupa jawaban siswa dan data yang dimasukkan guru berupa kunci jawaban buku yang dimasukkan oleh guru. Dan semua data yang sudah dimasukkan di simpan ke dalam basis data.

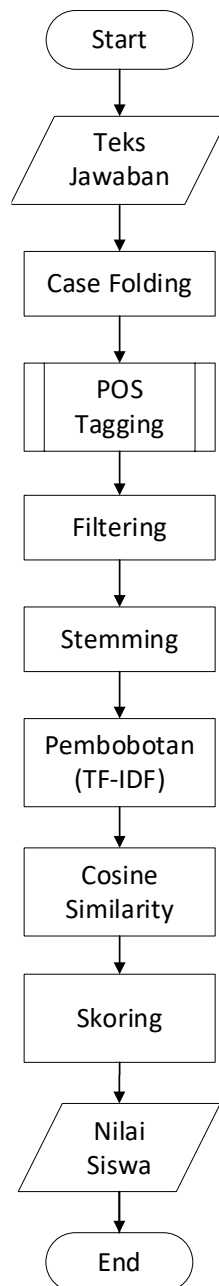
b. Proses

Data yang sudah dimasukkan kemudian di proses menggunakan *Case Folding* kemudian melakukan *POS Tagging* untuk memberi label pada setiap kata, selanjutnya melakukan *Filtering* dan *Stemming*, selanjutnya

melakukan pembobotan setiap *term* di dalam dokumen menggunakan *TF-IDF*, dan mengukur similaritas antara jawaban siswa dengan kunci jawaban menggunakan *Cosine Similarity*, dan skoring nilai dari hasil similaritas untuk setiap soal.

c. Output

Menghasilkan nilai ujian *essay* siswa.



Gambar 3. 2 Flowchart Sistem

Berikut merupakan alur atau tahapan dalam penilaian *essay* otomatis :

1. Input

Pada tahapan pertama memasukkan data berupa dokumen teks kunci jawaban guru dan jawaban siswa.

Jawaban / Q : bisa ular sangat berbahaya dan mematikan

D1 : Bisa ular sangat berbahaya dan mematikan

D2 : Bisa ular sangat mematikan.

D3 : Ular bisa sangat berbahaya.

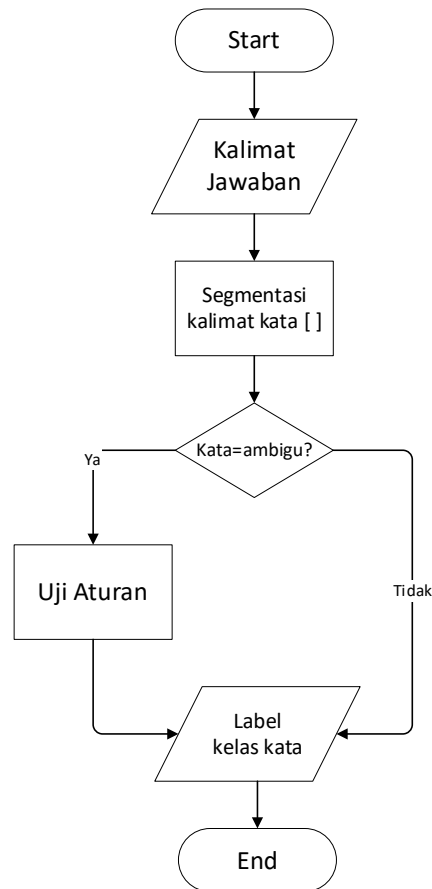
2. *Case Folding*

Pada tahapan ini mengubah teks menjadi *lowercase* atau huruf kecil, menghapus angka, menghapus tanda baca, dan menghapus *whitepace* atau karakter kosong.

Tabel 3. 1 Proses *Case Folding*

Data	Teks	<i>Case folding</i>
Q	bisa ular sangat berbahaya dan mematikan	bisa ular sangat berbahaya dan mematikan
D1	Bisa ular sangat berbahaya dan mematikan	bisa ular sangat berbahaya dan mematikan
D2	Bisa ular sangat mematikan.	bisa ular sangat mematikan
D3	Ular bisa sangat berbahaya.	ular bisa sangat berbahaya

3. POS Tagging metode Rule Based



Gambar 3. 3 Flowchart POS Tagging metode Rule Based

Deskripsi dari *flowchart* pada gambar 3.1 sebagai berikut:

- Memasukkan dokumen kalimat jawaban.
- Melakukan segmentasi kalimat dari dokumen menjadi per kata.
- Pengecekan ambiguitas kata dengan memeriksa pada setiap kata yang sudah disegmentasi apakah ada kata yang memiliki lebih dari satu kelas kata.
- Jika “tidak” akan melakukan maka melakukan pelabelan kata, Jika “iya” untuk kata yang memiliki lebih dari satu kata, akan dilakukan langkah selanjutnya.
- Uji dengan aturan yang sudah dibuat untuk menentukan kelas kata, gambarannya seperti kata “tahu” yang merupakan kata ambigu memiliki dua kelas kata yakni sebagai verbal dan nominal. Di dalam aturan Bahasa Indonesia, ciri kata nominal diikuti kata ini, itu, dan

tersebut. Berarti jika “tahu” diikuti kata itu, maka kata “tahu” memiliki kelas kata nominal bukan verbal. Dalam Langkah ini menghasilkan aturannya yaitu :

- Langkah 1 : mendeteksi kata sebelumnya.
- Langkah 2 : mendeteksi label kata sebelumnya.
- Langkah 3 : mendeteksi kata sesudahnya.
- Langkah 4 : mendeteksi label kata sesudahnya.

f. Terakhir menampilkan label *POS Tagging*.

Setelah melakukan segmentasi kalimat akan melakukan pengecekan ambiguitas kata dengan melalui uji aturan yang sudah ditetapkan jika kata memiliki lebih dari satu makna maka untuk menentukan posisi label atau tag dengan tepat untuk melakukan *POS Tagging*,

Pada jawaban tersebut terdapat satu kata yang bermakna lebih dari satu yaitu “bisa” pada Q, D1 dan D2 bermakna kata bisa merujuk pada zat racun dan di D3 bermakna kata bisa merujuk pada kata kerja yang berarti mampu atau dapat, berikut pada gambar 3.1 merupakan contoh pelabelan kata *POS Tagging*.

Contoh Ilustrasi Pelabelan



Gambar 3. 4 Contoh Proses Pelabelan *POS Tagging*

Berikut merupakan hasil *POS Tagging* menggunakan metode *Rule Based* pada tabel 3.2

Tabel 3. 2 Hasil *POS Tagging*

Data	<i>POS Tagging</i>
Q	("bisa","NNP"), ("ular","NN"), ("sangat""RB"), ("berbahaya","VB"), ("dan","CC"), ("mematikan","VB")
D1	("Bisa","NNP"), ("ular","NN"), ("sangat""RB"), ("berbahaya","VB"), ("dan","CC"), ("mematikan","VB")
D2	("Bisa","NNP"), ("ular","NN"), ("sangat""RB"), ("mematikan","VB")
D3	("Bisa","NNP"), ("ular","NN"), ("sangat""RB"), ("berbahaya","VB")

4. *Filtering*

Pada tahapan ini data dari proses *POS Tagging* akan diolah menjadi data terstruktur agar lebih muda dalam proses penilaian. Pada tahapan ini mengambil kata-kata penting atau menghilangkan kata yang tidak memiliki makna, contohnya “yang”, “dan”, “di”, “ke”, “dari”, dan lainnya.

Tabel 3. 3 Proses *Filtering*

Data	Teks	<i>Filtering</i>
Q	bisa(NNP) ular(NN) sangat(RB) berbahaya(VB) dan(CC) mematikan(VB)	bisa(NNP) ular(NN) sangat(RB) berbahaya(VB) mematikan(VB)
D1	bisa(NNP) ular(NN) sangat(RB) berbahaya(VB) dan(CC) mematikan(VB)	bisa(NNP) ular(NN) sangat(RB) berbahaya(VB) mematikan(VB)
D2	bisa(NNP) ular(NN) sangat(RB) mematikan(VB)	bisa(NNP) ular(NN) sangat(RB) mematikan(VB)
D3	ular(NN) bisa(MD) sangat(RB) berbahaya(VB)	ular(NN) bisa(MD) sangat(RB) berbahaya(VB)

5. Stemming

Pada tahapan ini menghilangkan *infleksi* setiap kata pada dokumen menjadi kata dasar.

Tabel 3. 4 Proses Stemming

[bisa] NNP	[ular] NN	[sangat] RB
[bahaya] VB	[mati] VB	[bisa] MD

6. Menghitung *TF-IDF* setiap kata yang sudah di dapatkan dari proses *preprocessing*.

Tabel 3. 5 Term Dokumen Matriks

Kata Dasar	Q	D1	D2	D3
bisa (NNP)	1	1	1	0
ular (NN)	1	1	1	1
sangat (RB)	1	1	1	1
bahaya (VB)	1	1	0	1
mati (VB)	1	1	1	0
Bisa (MD)	0	0	0	1

Pertama dengan mencari *IDF* setiap kata pada dokumen, contoh disini dengan menghitung W_{IDF} dari kata bisa(NNP) :

$$\begin{aligned}
 W_{IDF} &= 1 + \log \frac{3}{2} \\
 &= 1 + \log 1,5 \\
 &= 1 + 0,17609 \\
 &= 1,17609
 \end{aligned}$$

Sehingga diperoleh nilai W_{IDF} dari kata dalam *term* matriks pada tabel 3.6

Tabel 3. 6 Hasil Perhitungan *IDF*

bisa	ular	sangat	bahaya	mati	bisa
------	------	--------	--------	------	------

(NNP)	(NN)	(RB)	(VB)	(VB)	(MD)
1,17609	0	0	1,17609	1,17609	1,17609

Selanjutnya menghitung *TF-IDF* pada setiap kata pada setiap dokumen, contohnya disini dengan menghitung kata bisa(NNP) pada D1 atau document jawaban 1

$$\begin{aligned}
 W_{tf-idf}(t_i, d_j) &= W_{TF}(t_i, d_j) \times W_{IDF}(t_i, d_j) \\
 &= 1 \times 1,17609 \\
 &= 1,17609
 \end{aligned}$$

Dan didapatkan hasil perhitungan *TF-IDF* setiap kata pada setiap dokumen pada tabel 3.7

Tabel 3. 7 Hasil Perhitungan *TF-IDF*

Kata Dasar	Q	D1	D2	D3
bisa (NNP)	1,17609	1,17609	1,17609	0
ular (NN)	1	1	1	1
sangat (RB)	1	1	1	1
bahaya (VB)	1,17609	1,17609	0	1,17609
mati (VB)	1,17609	1,17609	1,17609	0
Bisa (MD)	0	0	0	1,47712

7. Menghitung *Cosine Similarity* setiap dokumen yang sudah di dapatkan dari hasil perhitungan *TF-IDF* pada tabel 3.8 . contoh perhitungan *Cosine Similarity* pada D1 atau dokumen jawaban 1

$$\begin{aligned}
 q_i \cdot d_i &= (1,17609 \times 1,17609) + (1 \times 1) + (1 \times 1) + \\
 &\quad (1,17609 \times 1,17609) + (1,17609 \times 1,17609) \\
 &= 6,14957
 \end{aligned}$$

$$\begin{aligned}
 |q_i| &= \sqrt{1,17609^2 + 1^2 + 1^2 + 1,17609^2 + 1,17609^2} \\
 &= 2,47983
 \end{aligned}$$

$$\begin{aligned}
 |d_i| &= \sqrt{1,17609^2 + 1^2 + 1^2 + 1,17609^2 + 1,17609^2} \\
 &= 2,47983
 \end{aligned}$$

Perhitungan *Cosine Similarity* :

$$\begin{aligned} \text{CosSim}(d_i, d_j) &= \frac{6,14957}{2,47983 \cdot 2,47983} \\ &= 1 \end{aligned}$$

Berdasarkan hasil perhitungan tersebut, maka jawaban D1 memiliki nilai kemiripan sebesar 1, dan untuk hasil perhitungan dokumen yang lain pada tabel 3.8 . Setelah ditemukan nilai kemiripan lalu masuk ke tahap skoring.

Tabel 3. 8 Hasil perhitungan *Cosine Similarity*

D1	1
D2	0,880384
D3	0,741721

8. Skoring (Nilai akhir)

Dari hasil pengukuran nilai kemiripan maka akan dikonversi ke nilai akhir. Maka rentang skala nilai kemiripan pada tabel 3.9

Tabel 3. 9 Skoring

Rentang nilai kemiripan	Nilai Akhir
0,01-0,05	5
0,06-0,10	10
0,11-0,15	15
0,16-0,20	20
0,21-0,25	25
0,26-0,30	30
0,31-0,35	35
0,36-0,40	40
0,41-0,45	45
0,46-0,50	50
0,51-0,55	55
0,56-0,60	60
0,61-0,65	65
0,66-0,70	70

0,71-0,75	75
0,76-0,80	80
0,81-0,85	85
0,86-0,90	90
0,91-0,95	95
0,96-1,00	100

Dari konversi nilai kemiripan pada tabel 3.10 maka dapat ditemukan nilai akhir jawaban siswa sebagai berikut :

$$D1 = 100$$

$$D2 = 90$$

$$D3 = 75$$

3.3 DataSet

Data yang digunakan dalam penelitian ini berupa teks atau kalimat berbahasa Indonesia, berupa soal *essay*, kunci jawaban, dan jawaban siswa. Penelitian ini dilakukan di SMPN 1 Kalianget pada siswa kelas IX-A yang berjumlah 19 siswa dengan 5 soal dan kunci jawaban mata pelajaran IPA pada bidang Biologi dari bank soal kurikulum 2013 dan buku IPA kelas IX PT Intan Pariwara, dan setiap soal mempunyai bobot yang sama yaitu 20%. Data Soal dan jawaban dari buku Intan Pariwara pada Tabel 3.10. Data tersebut dapat dilihat pada lampiran 1.

Tabel 3. 10 Soal dan jawaban dari buku Intan Pariwara

No	Soal	Jawaban
1	Bagaimana ciri-ciri biji yang penyebarannya dibantu oleh angin?	Biji yang persebarannya dibantu oleh angin memiliki ciri berukuran kecil ringan dan memiliki struktur yang menyerupai sayap sehingga mudah terbawa angin.
2	Mengapa kultur jaringan menggunakan bagian tumbuhan yang masih muda?	Karena sel-sel pada jaringan muda memiliki kemampuan pembelahan dan sifat totipotensi yang lebih baik dibandingkan sel-sel pada jaringan yang sudah dewasa.
3	Bagaimana proses inseminasi buatan dilakukan?	Inseminasi buatan dilakukan dengan memasukkan sperma pejantan ke saluran

		reproduksi induk betina dengan menggunakan alat suntik
4	Mengapa pada bagian tanaman yang dicangkok dapat muncul akar? Jelaskan!	Mencangkok dilakukan dengan mengelupas bagian kulit batang dan dibalut dengan tanah atau sekam, bagian kulit batang dihilangkan untuk mencegah kambium membentuk floem kembali. Apabila tidak terdapat floem pada bagian terkelupas tersebut maka zat makanan dari bagian atas sayatan tidak dapat dialirkan ke bawah. Hal tersebut mengakibatkan penumpukan auksin dan karbohidrat yang akan merangsang terbentuknya akar pada bagian yang terkelupas.
5	Mengapa penyerbukan bastar jarang terjadi secara alami?	penyerbukan bastar merupakan peristiwa menempelnya serbuk sari pada kepala putik tumbuhan lain yang berbeda jenis. serbuk sari yang berasal dari tumbuhan jenis lain tidak dapat menempel pada kepala putik dengan kuat sehingga mudah lepas. hal tersebut terjadi karena adanya gaya adhesi dan kohesi antar molekul yang berbeda. lepasnya serbuk sari menyebabkan tidak terbentuknya Buluh serbuk dan tidak terjadinya penyerbukan

Penelitian ini juga menggunakan data korpus bahasa Indonesia dari Fakultas Ilmu Komputer Universitas Indonesia atau [POS Tag - Corpus Site \(ui.ac.id\)](http://pos.tag-ui.ac.id) yang dapat di akses secara online dan gratis yang berada di bawah lisensi *Creative Common*, terdiri dari 10.000 kalimat yang dibangun dari 256683 token, dengan format menggunakan *tab separated value* yaitu setiap baris dari token dengan nilai postag yang dipisahkan oleh karakter tab.

Tabel 3. 11 Tagset POS Tagging

No	Tag	Keterangan	Contoh
1	CC	<i>Coordinating conjunction</i>	dan, tetapi, atau
2	CD	Cardinal number	dua, juta, enam, 7916, sepertiga, 0,025, 0,525, banyak, kedua, ribuan, 2007, 25
3	OD	Ordinal number	ketiga, ke-4, pertama
4	DT	Determiner / article	Para, Sang, Si

5	FW	Foreign word	Perubahan iklim, ketentuan dan kondisi
6	IN	Preposition	dalam, dengan, di, ke, oleh, pada, untuk
7	JJ	Adjective	bersih, panjang, hitam, lama, jauh, marah, suram, nasional, bulat
8	MD	Modal and auxiliary verb	boleh, harus, sudah, mesti, perlu
9	NEG	Negation	tidak, belum, jangan
10	NN	Noun	monyet, bawah, sekarang, rupiah
11	NNP	Proper noun	Boediono, Laut Jawa, Indonesia, India, Malaysia, Bank Mandiri, BBKP, Januari, Senin, Idul Fitri, Piala Dunia, Liga Primer
12	NND	Classifier, partitive, and measurement noun	orang, ton, helai, lembar
13	PR	Demonstrative pronoun	ini, itu, sini, situ
14	PRP	Personal pronoun	saya, kami, kita, kamu, kalian, dia, mereka
15	RB	Adverb	sangat, hanya, justru, niscaya, segera
16	RP	Particle	pun, -lah, -kah
17	SC	Subordinating conjunction	sejak, jika, seandainya, supaya, meski, seolah-olah, sebab, maka, tanpa, dengan, bahwa, yang, lebih ... daripada ..., semoga
18	SYM	Symbol	IDR, +, %, @
19	UH	Interjection	brengsek, oh, ooh, aduh, ayo, mari, hai
20	VB	Verb	merancang, mengatur, pergi, bekerja, tertidur
21	WH	Question	siapa, apa, mana, kenapa, kapan, di mana, bagaimana, berapa
22	X	Unknown	statemen
23	Y	Punctuation	"...", ?, .

3.4 Skenario Pengujian

Skenario pengujian ini dilakukan untuk mengetahui perbandingan akurasi terhadap kombinasi *POS Tagging* dan metode *Cosine Similarity* dalam penilaian jawaban *essay* otomatis pada soal Biologi. Pengukuran akurasi adalah hasil pengukuran yang ditujukan pada kedekatan dengan nilai asli atau sesungguhnya[13]. Menggunakan metode *Root Mean Square Error (RMSE)*

merupakan metode pengukuran dengan cara mengukur perbedaan nilai prediksi sebagai estimasi dari nilai yang diobservasi. Cara menghitung *RMSE* yaitu dengan mengurangi nilai prediksi dengan nilai sejati kemudian di kuadratkan selanjutnya hasilnya dijumlahkan dan dibagi dengan banyaknya data.

$$RMSE = \sqrt{\frac{\sum_{t=1}^n (A_t - F_t)^2}{n}} \quad \textbf{Rumus 3. 1 RMSE}$$

Dimana :

RMSE : *Root Mean Square Error*

A_t : Nilai prediksi

F_t : Nilai sejati

n : Jumlah data

Tabel 3. 12 Nilai sistem dan guru

	Nilai Sistem (Nilai prediksi)	Nilai Guru (Nilai Sejati)
D1	100	100
D2	90	85
D3	75	60

Selanjutnya dengan mencari *RMSE*

$$RMSE_{D1} = \sqrt{(100 - 100)^2} = 0$$

$$RMSE_{D2} = \sqrt{(90 - 85)^2} = 5$$

$$RMSE_{D3} = \sqrt{(75 - 60)^2} = 15$$

Tabel 3. 13 Hasil perhitungan *RMSE*

	RMSE
D1	0
D2	5
D3	15
Jumlah	20
Rata-rata	6,66667

Dan untuk mencari akurasinya dengan menghitung 100 dikuangi dengan nilai jumlah dari *RMSE*.

$$Akurasi = 100 - RMSE$$

Rumus 3. 2 Akurasi

$$Akurasi = 100 - 6,66667$$

$$= 93,3333\%$$

Sehingga menghasilkan akurasi 93,3333.

Setelah sistem selesai dibuat sistem perlu dilakukan sebuah evaluasi, skenario pengujian adalah beberapa pengujian yang dilakukan untuk mendapatkan nilai evaluasi, pada penelitian ini skenario pengujian dilakukan dengan cara mencari nilai *RMSE* dari sistem yang menggunakan *POS Tagging* dan sistem yang tidak menggunakan *POS Tagging*, berikut merupakan rincian pengujian.

Tabel 3. 12 Skenario Pengujian

No	Skenario pengujian
1	Sistem menggunakan <i>POS Tagging</i> dan <i>Cosine Similarity</i>
2	Sistem tanpa menggunakan <i>POS Tagging</i> dan menggunakan <i>Cosine Similarity</i>

BAB IV

HASIL DAN PEMBAHASAN

Pada bab ini akan dijelaskan tahapan - tahapan mulai dari tahapan Proses Pengembangan Sistem, Lingkungan Uji Coba, dan Hasil Skenario.

4.1 Proses Pengembangan Sistem

1. Load Data

```
cur = mysql.get_db().cursor()
cur.execute("select * from soal where no='"+str(soal)+"'")
data = cur.fetchall()
cur.close()

cur = mysql.get_db().cursor()
cur.execute("select j.*,s.nama from jawaban j join siswa s on (j.siswa=s.no) where j.soal='"+str(soal)+"' ")
jawabana = cur.fetchall()
cur.close()
```

Tahapan awal yaitu pengambilan data yang akan diolah diambil dari database, variabel *cur* digunakan untuk menghubungkan dengan database, untuk soal yang digunakan menggunakan *syntax* "select * from soal where no='"+str(soal)+"'" untuk jawaban dari siswa menggunakan *syntax* "select j.*,s.nama from jawaban j join siswa s on (j.siswa=s.no) where j.soal='"+str(soal)+"'".

2. Case Folding

```
for x in range(len(dokumenOlah)):
    dokumenOlah[x] = dokumenOlah[x].casefold()
def cleansing(query): #parameternya query = kata yang mau
    dicleaning
    res="" #kata kosong
    for x in range(len(query)): #loop kata perkata di
        query
        if query[x] not in string.punctuation: #jika
            katanya termasuk kota
```

```

        res+=query[x] #res += kata
    return res #return res

```

Tahapan selanjutnya setelah mengambil data yaitu melakukan *Case Folding* pada data yang digunakan *Case Folding* disini menggunakan fungsi di *Python* yaitu `casefold()`. Tahapan *Cleansing* yaitu tahapan pembersihan karakter pada data, tahapan ini dapat di lihat pada *syntax* `string.punctuation` dimana *syntax* tersebut akan memfilter karakter yang tidak diperlukan pada data.

3. POS Tagging

```

def postmanu(q,katacorpus,tagscorpus):
    hasilpost=[]
    for g in q.split(" "):
        pos = [i for i in range(len(katacorpus)) if
katacorpus[i]==g]
        temp=[]
        for x in range(len(pos)):
            if tagscorpus[pos[x]] not in temp:
                temp.append(tagscorpus[pos[x]])
            if temp==[]:
                temp=['X']
            hasilpost.append(temp)
        hasil=[]
        for x in range(len(hasilpost)):
            if len(hasilpost[x]) ==2:
                pos = [i for i in range(len(katacorpus)) if
katacorpus[i]==q.split(" ")[x]]
                for y in range(len(pos)):

                    if x==0:
                        if tagscorpus[pos[y]+1] ==
hasilpost[x+1][0] and q.split("
") [x+1]==katacorpus[pos[y]+1]:
                            hasilpost[x] =
[tagscorpus[pos[y]]]
                            break
                    elif x==len(hasilpost)-1:
                        if tagscorpus[pos[y]-1] ==
hasilpost[x-1][0] and q.split(" ")[x-
1]==katacorpus[pos[y]-1]:
                            hasilpost[x] =
[tagscorpus[pos[y]]]
                            break
                    else:
                        if tagscorpus[pos[y]-1] ==
hasilpost[x-1][0] and tagscorpus[pos[y]+1] ==
hasilpost[x+1][0]:
                            hasilpost[x] =
[tagscorpus[pos[y]]]
                            break

```

```

        if len(hasilpost[x]) == 2:
            hasilpost[x] = [hasilpost[x][0]]
        try:
            hasil.append((q.split("
") [x], hasilpost[x][0]))
        except:
            hasil.append((q.split(" ") [x], 'X'))
    return hasil

```

Tahapan *POS Tagging* yaitu memberi tag pada setiap kata, tag yang digunakan yaitu kamus korpus Bahasa Indonesia, untuk kata yang memiliki lebih satu tag akan di cek kembali terhadap kata sebelum dan sesudahnya dapat dilihat pada *syntax* `if tagscorpus[pos[y]-1] == hasilpost[x-1][0] and tagscorpus[pos[y]+1] == hasilpost[x+1][0]`.

4. Filtering

```

def filtering(stopword, query):
    hasil = stopwords.remove(query) #return hasil
    filteringnya
    return hasil

```

Tahapan *Filtering* yaitu tahapan untuk menghilangkan kata penghubung , dalam sistem ini *Filtering* menggunakan library *stopword* dimana dalam proses *Filtering* menggunakan *syntax* `stopword.remove`.

5. Stemming

```

def stemming(stemmer, query):
    return stemmer.stem(query)

```

Tahapan *Stemming* yaitu tahapan untuk mengubah kata menjadi kata dasar, dalam sistem ini *Stemming* menggunakan library *Stopword* dimana dalam proses *Stemming* menggunakan *syntax* `stemmer.stem`.

6. TF-IDF

```

tf=[]
for kata in semua_kata:
    temp=[]
    temp.append(query.count(kata))
    for jawaban in dokumenOlah:
        temp.append(jawaban.count(kata))
    tf.append(temp)

#IDF
idf=[]
for nilai_tf in tf:

```

```

df = sum(nilaitf) - nilaitf[0]
try:
    ndf = len(dokumenOlah)/df
    idf.append(math.log10(ndf))
except:
    idf.append(0)

#TFIDF
tfidf=[]
for x in range(len(tf)):
    temp=[]
    for y in range(len(tf[x])):
        temp.append(tf[x][y]*idf[x])
    tfidf.append(temp)

```

TF-IDF adalah tahapan pembobotan kata, *TF-IDF* adalah menghitung kemunculan kata pada suatu data, **for** kata **in** semua kata adalah perulangan untuk setiap kata yang ada pada semua dokumen, temp.append(query.count(kata)) yaitu menghitung kemunculan kata pada *query*, temp.append(dok.count(kata)) untuk perhitungan *IDF* terdapat pada *syntax* idf.append(math.log2(ndf)), sedangkan *TF-IDF* terdapat pada *syntax* temp.append(tf[x][y]*idf[x]).

7. Cosine Similarity

```

#AKARSIGMAW2
akarsigmaw2=[]
for y in range(len(tfidf[0])):
    hasil=0
    for x in range(len(tf)):
        hasil += (tfidf[x][y]**2)
    akarsigmaw2.append(hasil**0.5)

#WQWD
wqwd=[]
for y in range(1,len(tfidf[0])):
    hasil=0
    for x in range(len(tf)):
        hasil+=(tfidf[x][0]*tfidf[x][y])
    wqwd.append(hasil)

#COSINE
similarity=[]
for x in range(len(wqwd)):
    try:
        cosine = wqwd[x] /
        (akarsigmaw2[0]*akarsigmaw2[x+1])
    except:
        cosine=0
    if cosine>1:

```

```
cosine=1  
similarity.append(cosine)
```

Tahapan selanjutnya yaitu menghitung kemiripan menggunakan metode *Cosine Similarity*, hasil *Cosine Similarity* disimpan pada variabel *Cosine* yang mana nilai *Cosine* nantinya akan dijadikan dalam pemberian nilai pada jawaban siswa.

8. Skoring Penilaian

```
nilaisiswa=[]  
for x in range(len(dokumenOlah)) :  
    if similarity[x] >=0 and similarity[x] <=0.05:  
        nilai = 5  
    if similarity[x] >0.05 and similarity[x] <=0.1:  
        nilai = 10  
    if similarity[x] >0.1 and similarity[x] <=0.15:  
        nilai = 15  
    if similarity[x] >0.15 and similarity[x] <=0.2:  
        nilai = 20  
    if similarity[x] >0.2 and similarity[x] <=0.25:  
        nilai = 25  
    if similarity[x] >0.25 and similarity[x] <=0.30:  
        nilai = 30  
    if similarity[x] >0.3 and similarity[x] <=0.35:  
        nilai = 35  
    if similarity[x] >0.35 and similarity[x] <=0.4:  
        nilai = 40  
    if similarity[x] >0.4 and similarity[x] <=0.45:  
        nilai = 45  
    if similarity[x] >0.45 and similarity[x] <=0.5:  
        nilai = 50  
    if similarity[x] >0.5 and similarity[x] <=0.55:  
        nilai = 55  
    if similarity[x] >0.55 and similarity[x] <=0.6:  
        nilai = 60  
    if similarity[x] >0.6 and similarity[x] <=0.65:  
        nilai = 65  
    if similarity[x] >0.65 and similarity[x] <=0.7:  
        nilai = 70  
    if similarity[x] >0.7 and similarity[x] <=0.75:  
        nilai = 75  
    if similarity[x] >0.75 and similarity[x] <=0.8:  
        nilai = 80  
    if similarity[x] >0.8 and similarity[x] <=0.85:  
        nilai = 85  
    if similarity[x] >0.85 and similarity[x] <=0.9:  
        nilai = 90  
    if similarity[x] >=0.9 and similarity[x] <=0.95:  
        nilai = 95  
    if similarity[x] >0.95 and similarity[x] <=1:  
        nilai = 100
```



```
nilaisiswa.append(nilai)
```

Proses terakhir yaitu pemberian nilai terhadap jawaban siswa berdasarkan nilai *Cosine* tertentu dapat dilihat di code diatas untuk kondisi pemberian nilai terhadap siswa.

4.2 Lingkungan Uji Coba

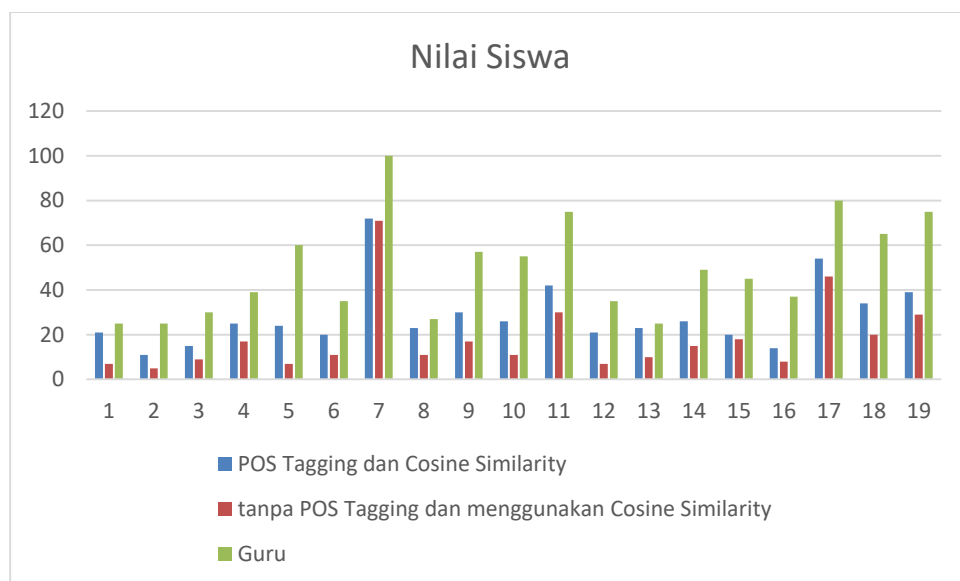
Pada sub bab berikut akan dijelaskan tentang spesifikasi perangkat lunak dan perangkat keras yang dipergunakan pada saat uji coba sistem.

Tabel 4. 1 Spesifikasi Perangkat Keras

No	Kebutuhan	Jenis
1	Sistem Operasi	<i>Windows 10</i>
2	<i>CPU</i>	Core i7 generasi 8
3	<i>RAM</i>	8 GB

4.3 Hasil Skenario

Penelitian ini dilakukan di SMPN 1 Kalianget pada siswa kelas IX-A yang berjumlah 19 siswa dengan 5 soal dan kunci jawaban mata pelajaran IPA pada bidang Biologi dengan perbandingan nilai menggunakan *POS Tagging* dan *Cosine Similarity*, dengan nilai tanpa *POS Tagging* dan menggunakan *Cosine Similarity*, dengan nilai dari Guru, dapat dilihat pada Gambar 4.1



Gambar 4. 1 Nilai Siswa

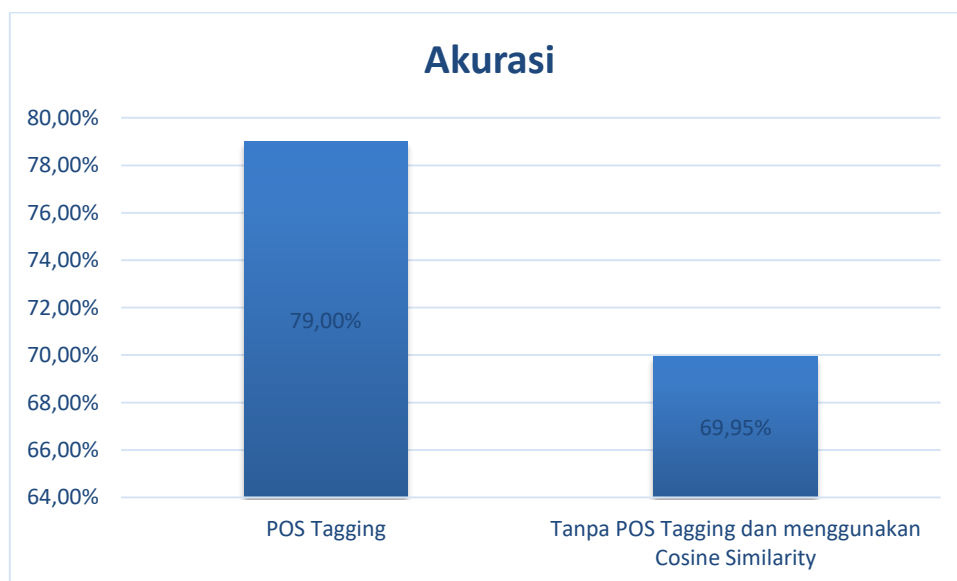
Pada grafik nilai siswa terlihat bahwa nilai guru lebih besar dari pada nilai sistem penilaian jawaban *essay* otomatis menggunakan *POS Tagging* dan *Cosine Similarity* dengan nilai sistem tanpa *POS Tagging* dan menggunakan *Cosine Similarity*, jadi nilai yang mendekati nilai guru yaitu nilai sistem yang menggunakan *POS Tagging* dan *Cosine Similarity*.

Setelah itu dilakukan skenario pengujian pada sistem, maka dihitung nilai *RMSE* dari pengujian sistem menggunakan kombinasi *POS Tagging* dan *Cosine Similarity*, dengan pengujian sistem tanpa menggunakan *POS Tagging* dan menggunakan *Cosine Similarity*. Berikut merupakan hasil pengujian pada tabel 4.2

Tabel 4. 2 Hasil Skenario Pengujian

No	Skenario Pengujian	<i>RMSE</i>	Akurasi
1	<i>POS Tagging</i> dan <i>Cosine Similarity</i>	21,0	79,0%
2	tanpa <i>POS Tagging</i> dan menggunakan <i>Cosine Similarity</i>	30,05	69,95%

Berikut merupakan hasil pengujian sistem menggunakan kombinasi *POS Tagging* dan *Cosine Similarity*, dengan pengujian sistem tanpa menggunakan *POS Tagging* dan menggunakan *Cosine Similarity* dengan bentuk grafik.



Gambar 4. 2 Grafik Akurasi

Terlihat bahwa pada grafik akurasi, nilai akurasi sistem penilaian jawaban *essay* otomatis menggunakan *POS Tagging* dan *Cosine Similarity* lebih besar dari nilai akurasi sistem tanpa *POS Tagging* dan menggunakan *Cosine Similarity*.

BAB V

PENUTUP

5.1 Kesimpulan

Berdasarkan hasil uji coba yang dilakukan pada sistem mendapat rata-rata akurasi dari setiap skenario pengujian yang dilakukan, sistem mendapatkan akurasi terbaik dalam penilaian jawaban *essay* otomatis menggunakan kombinasi *POS Tagging* dan metode *Cosine Similarity* dengan nilai akurasi 79,0%.

5.2 Saran

Penilaian jawaban *essay* otomatis menggunakan kombinasi *POS Tagging* dan metode *Cosine Similarity* pada penelitian ini masih memiliki akurasi yang belum begitu baik. Dan beberapa metode lain yang dapat dicoba seperti metode *POS Tagging* yang lain, seperti metode *Hidden Markov Model (HMM)* dan lain sebagainya yang mungkin dapat memiliki akurasi yang lebih baik. Selain itu sistem juga dapat menggunakan metode teks maining seperti *Dice Similarity*, *Generalized Latent Semantic Analysis GLSA*, *Latent Semantic Analysis LSA*, *Rabin Karp*, dan sebagainya. Dan sistem juga dapat mengimplementasikan *Query Expansion* untuk mencari sinonim kata, Untuk mendapatkan hasil akurasi yang bervariasi sehingga mendapatkan sistem yang lebih baik.

REFERENSI

- [1] E. L. Amalia, A. J. Jumadi, I. A. Mashudi, and D. W. Wibowo, “Analisis Metode Cosine Similarity Pada Aplikasi Ujian Online Otomatis (Studi Kasus JTI POLINEMA),” *J. Teknol. Inf. dan Ilmu Komput.*, vol. 8, no. 2, p. 343, 2021, doi: 10.25126/jtiik.2021824356.
- [2] F. Rahimi and A. N. Asyikin, “Aplikasi Penilaian Ujian Essay Otomatis Menggunakan Metode Cosine Similarity,” *Poros Tek.*, vol. 7, no. 2, pp. 88–94, 2015, [Online]. Available: <http://ejurnal.poliban.ac.id/index.php/porosteknik/article/view/218>
- [3] A. S. Yazid and A. Fatwanto, “Penentuan Kelas Kata Pada Part of Speech Tagging Kata Ambigu Bahasa Indonesia,” *JISKA (Jurnal Inform. Sunan Kalijaga)*, vol. 2, no. 3, p. 157, 2018, doi: 10.14421/jiska.2018.23-05.
- [4] A. Zaenal, *Evaluasi pembelajaran*. 2016.
- [5] A. Dinakaramani, F. Rashel, A. Luthfi, and R. Manurung, “Designing an Indonesian part of speech tagset and manually tagged Indonesian corpus,” *Proc. Int. Conf. Asian Lang. Process. 2014, IALP 2014*, pp. 66–69, 2014, doi: 10.1109/IALP.2014.6973519.
- [6] I. W. Soper, I. W. Mawa, I. W. Nardi, F. Pendidikan, and S. Ikip, “Hakekat , Penggolongan Dan Pembentukan Kata Bahasa Indonesia,” *Wacana*, vol. XX, no. 2, 2020.
- [7] N. Hidayat and L. Afuan, “Penilaian Ujian Otomatis untuk Soal Bertipe Essay pada PJJ APTIKOM menggunakan Cosine Similarity,” *Semin. Nas. APTIKOM*, pp. 259–271, 2019.
- [8] A. Bella, “Ambigu Adalah,” *Pak Dosen*, 2022. <https://pakdosen.co.id/ambigu-adalah/>
- [9] K. K. Purnamasari and I. S. Suwardi, “Rule-based Part of Speech Tagger for Indonesian Language,” *IOP Conf. Ser. Mater. Sci. Eng.*, vol. 407, no. 1,

2018, doi: 10.1088/1757-899X/407/1/012151.

- [10] F. Fataruba, “Penerapan Metode Cosine Similarity Untuk Pengecekan Kemiripan Jawaban Ujian Siswa,” *J. Mhs. Tek. Inform.*, vol. 2, no. 2, pp. 88–95, 2018.
- [11] M. Kamayani, “Perkembangan Part-of-Speech Tagger Bahasa Indonesia,” *J. Linguist. Komputasional*, vol. 2, no. 2, p. 34, 2019, doi: 10.26418/jlk.v2i2.20.
- [12] N. Vendyansyah and Y. A. Pranoto, “Perancangan dan Pembuatan Aplikasi untuk Mendeteksi Kemiripan Jawaban Menggunakan Cosine Similarity,” *J. Tek. (Jurnal Fak. Tek. Univ. Islam Lamongan)*, vol. 13, no. 1, pp. 23–28, 2021.
- [13] Ruslan, Gunawan, and T. Suhatati, “Sistem Penilaian Otomatis Jawaban Esai Menggunakan Metode GLSA,” *Semin. Nas. Apl. Teknol. Inf. 2018*, pp. 6–15, 2018.

LAMPIRAN

DATA PENELITIAN MATA PELAJARAN IPA KELAS IX-A SMPN 1 KALIANGET

DATA SOAL DAN KUNCI JAWABAN

No	Soal	Jawaban
1	Bagaimana ciri-ciri biji yang penyebarannya dibantu oleh angin?	Biji yang persebarannya dibantu oleh angin memiliki ciri berukuran kecil ringan dan memiliki struktur yang menyerupai sayap sehingga mudah terbawa angin.
2	Mengapa kultur jaringan menggunakan bagian tumbuhan yang masih muda?	Karena sel-sel pada jaringan muda memiliki kemampuan pembelahan dan sifat totipotensi yang lebih baik dibandingkan sel-sel pada jaringan yang sudah dewasa.
3	Bagaimana proses inseminasi buatan dilakukan?	Inseminasi buatan dilakukan dengan memasukkan sperma pejantan ke saluran reproduksi induk betina dengan menggunakan alat suntik
4	Mengapa pada bagian tanaman yang dicangkok dapat muncul akar? Jelaskan!	Mencangkok dilakukan dengan mengelupas bagian kulit batang dan dibalut dengan tanah atau sekam, bagian kulit batang dihilangkan untuk mencegah kambium membentuk floem kembali. Apabila tidak terdapat floem pada bagian terkelupas tersebut maka zat makanan dari bagian atas sayatan tidak dapat dialirkan ke bawah. Hal tersebut mengakibatkan penumpukan auksin dan karbohidrat yang akan merangsang terbentuknya akar pada bagian yang terkelupas.
5	Mengapa penyerbukan bastar jarang terjadi secara alami?	penyerbukan bastar merupakan peristiwa menempelnya serbuk sari pada kepala putik tumbuhan lain yang berbeda jenis. serbuk sari yang berasal dari tumbuhan jenis lain tidak dapat menempel pada kepala putik dengan kuat sehingga mudah lepas . hal tersebut terjadi karena adanya gaya adhesi dan kohesi antar molekul yang berbeda. lepasnya serbuk sari menyebabkan tidak terbentuknya Buluh serbuk dan tidak terjadinya penyerbukan

DATA JAWABAN SISWA

No	NIS	Nama Siswa	L/P	Jawaban 1	Jawaban 2	Jawaban 3	Jawaban 4	Jawaban 5	Nilai
1	0074981885	Alfrian Suci Andani	P	Biji yang penyebarannya dibantu angin biasanya bentuknya lebih kecil dan warnanya lebih terang	Karena biasanya tumbuhan yang masih muda lebih gampang untuk dijadikan sebagai kultur jaringan		Karena batang tumbuhannya telah berada di dalam tanah, jadi jaringan yang berada di tumbuhan memunculkannya akar	Karena Serbuk sari berkemungkinan kecil bisa menempel di putik tumbuhan yang lain dengan sendirinya, jadi penyerbukannya jarang terjadi secara alami	25
2	0072111226	Anggun Febriyanti	P	Putik, berbunga	Karena lebih mudah untuk penyerbukannya.	Diambil sel spermanya.	Karena pada saat diberi sabut kelapa dan tanah batang akan tumbuh akar.	Karena bagian tumbuhan yang sudah tua lebih susah penyerbukannya.	25
3	0077448177	Berliana Cinta Tiarani	P	ringan dan punya sayap = contoh - dandelion, biasanya	Kultur jaringan : metode perbanyak tumbuhan melalui	Inseminasi : proses perkembangbiakan pada hewan dengan	Cangkok : mengelupas kulit suatu tanaman berkayu lalu	Karena penyerbukan bastar terjadi karena adanya, anemogami :)	30

				tumbuhannya banyak, mengandung nektar dan tidak bau.	pengambilan sel atau organ jika kultur jaringan menggunakan bagian tumbuhan yang masih muda biasanya akan cepat berbuah.	menyatukan sel sperma jantan ke sel betina. Proses : menyuntikkan sel sperma sapi jantan ke sapi betina.	dibungkus menggunakan sabut kelapa atau plastik. Kenapa dapat muncul akar : pencangkokan pada suatu tanaman akan menghasilkan tanaman jenis yang sama maka akibatnya akar baru dapat muncul.		
4	0073163 925	Deta Lailatul Qodriyah	P	Ciri-cirinya adalah tanaman berukuran kecil, bersifat ringan	Karena jika menggunakan yang masih muda memperbanyak tanaman yang tumbuh	Dengan cara penyuntikan sel sperma hewan jantan disuntikkan ke hewan betina	Karena batang tanaman berkayu dikelupas, kemudian dibalut dengan tanah dan dibungkus dengan sabut kelapa atau plastik sehingga	Karena penyerbukan bastar menempelnya benang sari ke kepala putik dari bunga pada tumbuhan lain yang berbeda jenis	39

							tumbuh akar		
5	0074740 690	Farelliana Rusti Nur Febriyani	P	Kecil, ringan, bersayap, tidak terlalu dipengaruhi oleh gravitasi bumi sehingga mudah terbawa angin. (anemokori)	Karena bagian tumbuhan yang masih muda, kebanyakan kandungan dan nutrisinya masih berfungsi dengan baik sehingga bisa tumbuh di tanaman lain.	Inseminasi buatan adalah teknologi perkembangbia kan pada hewan, dengan cara menyuntikkan sperma (jantan) ke dalam hewan betina, contohnya penyuntikan sel sperma sapi jantan kedalam sapi betina.	Mencangkok dilakukan dengan mengelupas kulit tanaman berkayu. lalu dibalut dengan tanah, dan dibungkus dengan sabut kelapa / plastik sehingga muncul akar baru. akar muncul karena bagian yang dicangkok mendapat kandungan nutrisi dari tanaman berkayu tadi sehingga dapat tumbuh.	Penyerbukan bastar terjadi pada dua tanaman bunga yang berlain- lainan jenis. Benang sari jatuh ke kepala putik dari tanaman lain. Benang sari terbawa angin dan mengikuti arah gerak angin sehingga jatuhnya tidak selalu pada kepala putik, dan memungkinan penyerbukan bastar jarang terjadi karena terbangnya benang sari terpaku pada arah geraknya	60

								angin.	
6	0062960 198	Hafizah Inayah sya'bania	P	bunga berukuran kecil, serbuk sari banyak dan ringan, jumlah bunga banyak, tidak menghasilkan bau / nektar	Karena suatu metode perbanyak tumbuhan dengan cara mengambil suatu bagian dari tumbuhan seperti sel atau sekelompok sel, jaringan, atau organ. Selanjutnya ditumbuhkan ke dalam kondisi steril pada medium yang mengandung nutrisi dan zat pengatur tubuh (hormon)	merangsang ovarium dengan memberikan hormon, membantu folikel ovarium secara berkala, menginduksi ovulasi dengan suntikan HCG minimal 1 folikel, menyiapkan sperma yang disiapkan dalam wadah, sperma dicuci secara khusus, melakukan inseminasi ketika betina berovulasi, memberikan progesteron	Karena aliran zat makanan dan auksin (hormon tumbuh yang mendorong keluarnya akar) dan terjadi penumbuhan karbohidrat dan yang mendorong keluarnya akar	Karena serbuk sari yang menempel pada kepala putik berasal dari bunga pada tumbuhan lain yang berbeda jenis atau sebaliknya memiliki satu sifat beda	35
7	0077764	Holila	P	Ciri biji yang	Karena sel-sel	Proses	Karena	Karena	100

	691	penyebarannya dibantu oleh angin : Biji berukuran kecil ringan, Struktur yang menyerupai sayap sehingga mudah terbawa angin.	yang ada pada jaringan muda memiliki kemampuan pembelahan dan sifat totipotensi yang lebih baik jika dibandingkan dengan sel-sel pada jaringan yang sudah dewasa.	inseminasi buatan dilakukan dengan memasukkan sperma hewan jantan ke saluran reproduksi induk hewan betina menggunakan alat suntik	mencangkok dilakukan bagian kulit batang dikelupas lalu dibalut dengan tanah dan dililitkan plastik, sehingga kulit batang yang sudah dihilangkan mencegah kambium membentuk floem kembali, karena jika terdapat floem pada bagian yang terkelupas maka zat makanan tidak dapat dialirkan dari bagian sayatan ke bawah, dan	penyerbukan bastar menempelnya serbuk sari pada kepala putik ke tumbuhan lain yang jenisnya berbeda, serbuk sari dari tanaman lain tidak dapat menempel dengan kuat pada kepala putik sehingga mudah terlepas, karena adanya gaya adesi dan kohesi antar molekul yang beda, karena serbuk sari terlepas menyebabkan tidak terbentuk buluh serbuk
--	-----	--	---	--	---	--

							mengakibatkan penumpukan auksin dan karbohidrat yang akan merangsang dan terbentuklah akar pada bagian yang terkelupas.	sehingga tidak terjadi penyerbukan.	
8	0066319773	Irma Nur Fadila	P	tidak menghasilkan bau, serbuk sari banyak dan ringan, jumlah bunga banyak	agar kultur jaringan dapat berlangsung dengan	merangsang ovarium dengan memberikan hormon, membantu partikel ovarium secara berkala, menginduksi ovulasi	karena pada batang terdapat dan ditempel dengan tanah	karena serbuk sari yang menempel pada kepala putik berasal dari bunga pada tumbuhan lain	27
9	0062405522	Kadaryono	L	kecil, ringan, bersayap tidak terlalu dipengaruhi gravitasi bumi sehingga mudah	Karena bagian tumbuhan yang masih muda, kebanyakan kandungan dan	Inseminasi buatan dilakukan dengan mengupas kulit batang dan	Mencangkok dilakukan dengan mengelupas bagian kulit batang atau	Karena serbuk sari pada putik berasal dari bunga pada tumbuhan lain.	57

				terbawa angin	nutrisinya masih berfungsi dengan baik sehingga bisa tumbuh di tanaman lain.	dibaluti dengan tanah atau sekam, lalu muncul akar pada bagian tersebut.	sekam lalu muncul akar.		
10	0079893663	Laila Andreyani Setia Diningrum	P	berukuran kecil, ringan, dan bersayap.	Karena dengan bagian tumbuhan yang masih muda dapat memperbanyak tanaman dengan cepat.	Inseminasi yaitu proses perkembangbiakan pada hewan, prosesnya dengan cara menyuntikkan sel sperma pada jantan ke sel telur betina. Sehingga, sel telur betina mengalami pembuahan.	karena batang berkayu yang sudah dikupas kulitnya, dibaluti dengan tanah dan dibungkus dengan sabut kelapa atau plastik yang menyebabkan bagian tanaman yang dicangkok itu muncul akar.	Karena suatu tanaman jarang berdekatan dengan tanaman lain yang berbeda jenis. Sehingga serbuk sari sulit untuk menempel pada kepala putik. Oleh karena itu, jarang terjadi penyerbukan bastar secara alami.	55
11	0078074013	M. Mustofa	L	berukuran kecil, ringan dan menyerupai sayap	Karena sel-sel pada jaringan muda memiliki kemampuan	Inseminasi yaitu proses perkembangbiakan pada	Karena batang berkayu sudah dikupas kulitnya,	karena suatu tanaman jarang berdekatan dengan	75

					pembelahan dan sifat totipotensi yang lebih baik jika dibandingkan dengan sel-sel pada jaringan dewasa.	hewan prosesnya dengan cara sel sperma pada jantan dimasukkan ke sel telur betina dengan alat suntik, sehingga sel telur mengalami pembuahan.	dibaluti dengan tanah dan dibungkus dengan sabut kelapa atau plastik yang menyebabkan bagian tanaman yang dicangkok itu muncul akar	tanaman lain yang berbeda jenis, sehingga serbuk sari sulit menempel dengan kepala putik. Oleh karena itu, jarang terjadi penyerbukan bastar secara alami	
12	0062994493	Maulana Alief Hidayatullah	L	Ciri biji yang penyerbukannya dibantu oleh angin kecil, tipis dan yang mempunyai sayap	Karena kultur jaringan yang menggunakan bagian tumbuhan yang masih muda karena masih sangat bagus dan masih sangat alami	Inseminasi buatan dilakukan dengan cara menempatkan sperma kepada rahim	Karena pada saat mencangkok tanaman pada tumbuhan kita harus menggali tanah / atau mengambil tanah lalu dicabut tanahnya dan muncullah akar	Karena dilakukan secara alami sangatlah susah, dan tidak memiliki mahkota bunga dan tangkai bunganya	35
13	0073886586	Mohammad Abdillah	L	Ciri-ciri biji yang	Karena sel-sel organ pada	Yaitu dengan memasukkan	Karena bagian yang sudah ter	Karena tidak menghasilkan	25

				penyebarannya dibantu oleh angin yaitu mempunyai sayap	tumbuhan yang masih muda masih bagus sedangkan kalau tumbuhan yang sudah tua sel-sel organnya sudah tidak bagus.	sperma hewan jantan ke dalam tabung, lalu dimasukkan ke dalam alat kelamin hewan betina dengan menggunakan alat	cangkok jika ditambah tanah dan dibungkus plastik lalu diikat, lalu dikasih air akan tumbuh akar.	nektar atau bau, tidak memiliki mahkota bunga, dan tangkai bunganya tidak lentur	
14	0074874 163	Nurmala barkiyah	P	berbiji kecil, ringan, dan bersayap. Contohnya : biji bunga dandelion	Kultur jaringan tumbuhan dilakukan dengan mengambil salah satu bagian tumbuhan seperti sel, kelompok sel, jaringan/ organ. kemudian bagian tumbuhan tersebut	dengan cara menyuntikkan sperma (semen) induk jantan yang unggul ke saluran perkembangbiakan induk betina. sperma (semen) biasanya dibekukan baru disuntikkan kepada induk	Cangkok dilakukan dengan mengelupas kulit tanaman berkayu kemudian bagian yang terkelupas ditemplei tanah. lalu ditutup dengan serabut kelapa / plastik, bagian tanaman yang	penyerbukan bastar terjadi jika benang sari yang menempel pada putik berada pada bunga tanaman lain yang berbeda jenis / setidaknya memiliki 1 sifat berbeda. Mengapa jarang terjadi secara alami?	49

					dikembangbiakan / ditumbuhkan pada tempat yang steril dan dibantu dengan hormon penumbuh tanaman. lalu bagian tumbuhan akan tumbuh daun dan organ tanaman yang lain, menggunakan bagian tumbuhan yang masih muda dilakukan agar mudah kultur jaringan tumbuhan	betina	dikelupas tersebut akan muncul akar setelah beberapa lama mengapa ? karena terdapat media tanah yang memungkinkan akar tumbuh dari bagian tanaman yang dikelupas	karena benang sari dari bunga lain dan putiknya juga dari tanaman lain sulit dikembangbiakan, seringkali benang sari tidak jatuh di atas kepala putik, dan disulitkan juga karena letaknya yang jauh antara putik dan benang sari 49	
15	0061398795	Rafli Sanjaya	L	ciri-cirinya adalah bunga berukuran kecil, serbuk sari	Karena kultur adalah suatu metode untuk mengisolasi	Proses inseminasi dilakukan dengan cara	Karena aliran zat makanan (karbohidrat) dan auksin	Karena serbuk sari yang berasal dari tumbuhan lain	45

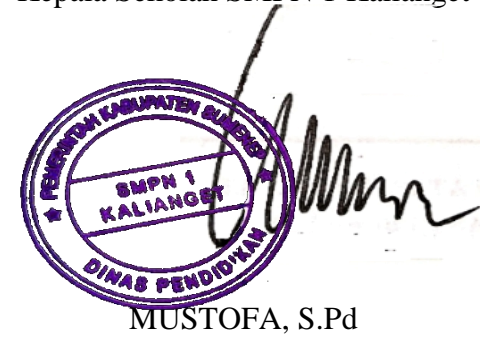
				banyak dan ringan, jumlah bunga banyak dan tidak menghasilkan bau atau nektar	bagian dari tanaman seperti sekelompok jaringan atau sel yang ditumbuhkan dalam kondisi aseptik sehingga bagian tanaman tersebut bisa memperbanyak diri menjadi tanaman baru kembali dengan sifat yang sama	menempatkan sperma langsung ke dalam rahim pada saat pelepasan sel telur (ovulasi) menggunakan kateter kecil.	(hormon tumbuh yang mendorong keluarnya akar) dan terjadi penimbunan karbohidrat dan hormon jadi meningkat dan berbentuk katus yang berubah menjadi akar tanaman	tidak dapat melekat kuat pada kepala putik, hal ini terjadi karena adanya gerakan adhesi dan kohesi	
16	0063702 848	Revaliza Ajaz Indra Rukmana	L	berbiji kecil, ringan, dan bersayap.	karena jika menggunakan yang tua tidak akan membuahkan hasil.	dengan cara menyuntikkan sel sperma jantan ke sel telur betina.	karena batang yang dikupas dibaluti dengan tanah dan ditutupi dengan sabut kelapa, atau plastik sehingga	karena jika terjadi secara buatan tidak akan terjadi pembuahan.	37

							muncullah akar.		
17	0066899 164	Rifka ayu Lestari	P	Ciri biji yang penyebarannya dibantu oleh angin adalah memiliki ukuran kecil dan ringan, dan strukturnya menyerupai sayap sehingga biji mudah terbawa angin.	Karena sel-sel pada jaringan yang masih muda memiliki kemampuan pembelahan dan sifat totipotensi yang lebih baik jika dibandingkan pada sel-sel jaringan yang sudah tua.	Inseminasi buatan dilakukan dengan menggunakan alat untuk memasukkan sperma jantan ke saluran reproduksi induk betina.	mencangkok dengan cara mengelupas bagian kulit batang dan dibalut dengan tanah, lalu muncul akar pada bagian tersebut.	Karena penyerbukan bastar menempelnya benang sari ke kepala putik dari bunga pada tumbuhan lain yang beda jenis.	80
18	0071469 453	Ringga Ali Fikri Zamani	L	tanaman yang penyebaran bijinya dibantu oleh angin (anemokori) memiliki ciri-ciri yakni : memiliki biji, yang kecil dan ringan serta memiliki sayap	Karena, tumbuhan yang masih muda masih dapat tumbuh dan berkembang jauh lebih baik dan cepat dibandingkan dengan	Proses inseminasi buatan dilakukan dengan cara menyuntikkan sel sperma kedalam sel telur, sehingga induk betina akan	tanaman yang dicangkok telah dibaluti dengan tanah dan dibungkus menggunakan sabut kelapa atau plastik dengan rapat, hingga udara tidak dapat	Penyerbukan bastar secara alami jarang terjadi karena serbuk sari tidak jatuh tepat di kepala putik dan bahkan dapat terbang jauh menjauhi	65

				sehingga mudah terbawa angin.	tanaman yang sudah tua.	mengalami proses pembuahan.	masuk, tanaman yang dicangkok dapat memperoleh zat hara yang cukup sehingga memungkinkan bagian tanaman yang dikelupas kulitnya dapat tumbuh akar.	kepala putik sehingga kecil kemungkinan untuk berhasil.	
19	0063883 160	Shofwan Dani Yulianto Adilah	L	biji yang persebarannya dibantu oleh angin : biji berukuran kecil dan ringan, struktur biji menyerupai sayap sehingga mudah terbawa angin	karena sel-sel yang ada pada jaringan muda memiliki kemampuan dan sifat totipotensi yang lebih baik dari sel-sel yang sudah dewasa.	proses inseminasi buatan dilakukan dengan menyuntikkan atau memasukkan sperma hewan jantan ke saluran induk hewan betina menggunakan alat suntik.	Karena batang yang kupas kulitnya dibaluti dengan tanah dan ditutup dengan tanah dan plastik sampai tumbuh akar.	Karena jika terjadi secara buatan tidak akan tumbuh	75

Mengetahui,

Kepala Sekolah SMPN 1 Kalianget

A handwritten signature in black ink is written over a circular official stamp. The stamp is purple and contains the text "PEMERINTAH KABUPATEN SUMBAWA" at the top, "SMPN 1 KALIANGET" in the center, and "DINAS PENDIDIKAN" at the bottom. The signature is a cursive script that starts with a long vertical stroke and ends with a series of loops.

MUSTOFA, S.Pd

NI P. 19650616 198911 1 001