

BAB I

PENDAHULUAN

I.1 Latar Belakang

Hasil Sensus Penduduk pada tahun 2010 menunjukkan dari berbagai bahasa daerah yang terdapat di Indonesia, Persentase penggunaan Bahasa Melayu yang digunakan oleh masyarakat di Kalimantan Barat mencapai 1.615.978 juta jiwa dari total penduduk Kalimantan Barat (Badan Pusat Statistik, 2011). Bahasa Melayu Pontianak merupakan dialek bahasa Melayu yang dituturkan oleh masyarakat Kota Pontianak, Kabupaten Kubu Raya dan Kabupaten Mempawah serta memiliki kesamaan dengan Bahasa Melayu Semenanjung Malaysia (Johor-Riau) (Putra, 2018). Upaya pelestarian bahasa Melayu Pontianak agar tidak punah dan ditinggalkan karena pengaruh globalisasi tetap harus dilakukan. Salah satu upaya dengan penerapan teknologi *Text To Speech* (TTS).

Teknologi *Text To Speech* (TTS) adalah suatu sistem yang dapat mengkonversi teks menjadi ucapan. Kualitas ucapan pada TTS dapat dinilai dari kejelasan dan kealamian ucapan yang dihasilkan. Indikator ucapan yang jelas dan alami adalah nada, tekanan (keras-lemah), durasi, intonasi, dan jeda (kesenyapan). Pemberian jeda dalam kalimat sangat penting karena jeda yang terjadi dapat memperjelas informasi dari makna atau maksud pada suatu teks kalimat dapat tersampaikan dengan benar. Sebagaimana fungsi tanda baca yang membantu memahami arti dari suatu teks, jeda juga meningkatkan kealamian dan kejelasan pengucapan bacaan (Venkatesh, 2007). Penelitian pernah dilakukan oleh M. Iqbal Kamiludin (2017) tentang Prediksi Jeda Dalam Ucapan Kalimat Bahasa Melayu Pontianak Menggunakan Metode *Shallow Parsing* tidak sepenuhnya mampu melakukan proses pemrediksian jeda pada Kalimat Bahasa Melayu Pontianak secara akurat karena total hasil perhitungan *F-measure* sebesar 0.64. Nilai *recall* dan *precision* untuk kalimat tunggal sebesar 0.78 dan 0.74 , sedangkan untuk kalimat mejemuk sebesar 0.67 dan 0.57. Potongan frasa jeda *shallow parsing* belum mampu menghasilkan nilai akurasi kecocokan pada frasa jeda penutur dengan nilai baik. Dari 168 kalimat yang ada, nilai kecocokan dengan jeda penutur sebesar 40.4% atau 68 kalimat.

Pemberian prediksi jeda bisa juga dilakukan dengan mengambil informasi parafrasa. Parafrasa yaitu ungkapan kembali maksud atau isi tulisan dengan menggunakan kata-kata sendiri (Sabarti,1996). Jadi, parafrasa yaitu mengungkapkan kembali berbagai informasi yang dibaca atau yang didengar dengan cara dan bentuk yang berbeda. Selain itu, parafrasa juga digunakan untuk menguraikan suatu teks dalam bentuk atau susunan kata yang lain agar makna yang tersembunyi dalam teks tersebut dapat dijelaskan. Diharapkan dengan menggunakan informasi parafrasa dapat menghasilkan prediksi jeda dengan lebih baik dan dengan proses yang lebih mudah.

Berdasarkan penjelasan yang dipaparkan, maka dilakukan penelitian untuk prediksi jeda dalam ucapan kalimat Bahasa Melayu Pontianak menggunakan informasi parafrasa agar dapat digunakan oleh sistem *Text To Speech* (TTS).

I.2 Perumusan Masalah

Berdasarkan latar belakang yang telah diatas, maka rumusan masalah yang muncul dalam penelitian ini adalah bagaimana mengambil informasi parafrasa dalam suatu teks kalimat yang digunakan untuk memprediksi jeda.

Informasi parafrasa dapat berupa kata kata tertentu, penekanan dalam suatu kata atau kalimat, tinggi rendahnya nada dalam suatu kalimat, dan panjang pendeknya durasi kalimat.

Pada penelitian ini, informasi parafrasa diambil berdasarkan kata kunci dan tipe PoS, dimana kejadian jeda yang terjadi bisa terdapat pada sebelum atau sesudah kata kunci. Kata kunci adalah kata yang akan ditetapkan dan digunakan dalam prediksi jeda dengan menggunakan tipe PoS.

I.3 Tujuan Penelitian

Adapun tujuan dari penelitian yang dilakukan adalah menghasilkan prediksi jeda dalam bahasa Melayu Pontianak menggunakan informasi parafrasa.

I.4 Pembatasan Masalah

Pembatasan masalah dari penelitian yang akan dilakukan adalah :

1. Korpus yang digunakan adalah korpus Bahasa Melayu Pontianak yang diambil dari buku Sepok Satu.
2. Data jeda diperoleh dari data jeda ucapan ucapan yang direkam dari oleh satu penutur ahli Bahasa Melayu Pontianak.

I.5 Sistematika Penulisan

Sistematika dari penulisan tugas akhir ini disusun dalam 5 (lima) bab yang terdiri dari Bab I Pendahuluan, Bab II Tinjauan Pustaka, Bab III Metodologi Penelitian dan Arsitektur Sistem, Bab IV hasil dan analisis serta Bab V penutup.

Bab I pendahuluan adalah bab yang berisi latar belakang, perumusan masalah, tujuan penelitian, pembatasan masalah, dan sistematika penulisan.

Bab II tinjauan pustaka adalah bab yang berisi uraian sistematis tentang informasi parafrasa, part of speech tagging, korpus, Bahasa Melayu Pontianak, precision & recall, F- measure, dan kajian terkait.

Bab III metodologi penelitian adalah bab yang berisi tentang data dan perangkat penelitian, metode penelitian, diagram alir penelitian dan analisa pengujian penelitian.

Bab IV hasil dan analisis adalah bab yang berisi tahapan pengujian, hasil pengujian dan analisis pengujian yang mengarah kepada suatu kesimpulan.

Bab V penutup adalah bab yang berisi kesimpulan dari penelitian telah dilakukan dan saran / rekomendasi untuk perbaikan, pengembangan atau kesempurnaan / kelengkapan penelitian yang telah dilakukan.

BAB II

TINJAUAN PUSTAKA

2.1 Kajian Terkait

Berikut adalah kajian terkait dari penelitian yang telah dibuat:

Prediksi Jeda Dalam Ucapan Kalimat Bahasa Indonesia Dengan *Hidden Markov Model* oleh Aditya Teguh Nugraha (2014) pada penelitiannya menjelaskan tentang perancangan aplikasi pemrediksi jeda dari ucapan kalimat Bahasa Indonesia dengan *Hidden Markov Model* (HMM). Aplikasi yang dirancang berfungsi untuk mengetahui kejadian jeda yang terjadi pada kalimat Bahasa Indonesia dengan dua proses utama yaitu, train untuk melatih korpus, dan prediksi untuk memprediksi jeda. Hasil dari penelitian tersebut adalah kalimat yang telah diprediksi kejadian jedanya dengan nilai *precision* sebesar 0.364, *recall* sebesar 0.132, dan *f-score* sebesar 0.194.

Prediksi Jeda Dalam Ucapan Bahasa Melayu Pontianak Dengan Menggunakan Metode *Shallow Parsing* oleh M. Iqbal Kamaludin (2017) pada penelitian nya menjelaskan tentang prediksi jeda pada ucapan Bahasa Melayu Pontianak dengan menggunakan metode *Shallow Parsing*. Penelitian ini belum mampu menghasilkan nilai akurasi kecocokan pada frasa jeda penutur dengan nilai yang baik. Dari 168 kalimat yang merupakan gabungan dari kalimat tunggal dan kalimat majemuk menghasilkan nilai kecocokan dengan jeda penutru sebesar 40.4% atau 68 kalimat. Hasil dari penelitian tersebut menghasilkan nilai recall 0.78 untuk kalimat tunggal 0.67 untuk kalimat majemuk yang menunjukkan bahwa sebagian besar frasa jed yang relevan tidak terbentuk oleh *shallow parsing*. Nilai precision 0.74 untuk kalimat tunggal dan 0.57 untuk kalimat majemuk yang menunjukkan nilai yang cukup baik namun masih banyak terdapat hasil potongan frasa jeda yang lain selain hasil potongan frasa jeda yang relevan yang ikut berbentuk *shallow parsing*.

Data-Driven Phrase Break Prediction for Bengali Text-to-Speech System oleh Krishnendu Ghosh dan K. Sreenivasa Rao (2012) menjelaskan tentang

prediksi frasa dengan menggunakan bahasa Bengali India dengan membandingkan pendekatan *Feed-Forward Neural Network* (FFNN) dan *Classification and Regression Tree* (CART) model. Korpus yang digunakan yaitu sebanyak 6.210 kalimat bahasa Bengali India. Salah satu pendekatan prediksi frasa berbasis aturan yang paling awal pada penelitian ini difokuskan pada struktur sintaksis dan organisasi psikologis dari suatu kalimat. Organisasi psikologis dari suatu kalimat adalah membagi sebuah kalimat dengan mempertimbangkan di mana manusia cenderung memperkenalkan jeda sambil mengucapkan kalimat itu. Berdasarkan analisis, tiga kategori berbeda (mis. jeda pendek, sedang, dan panjang) dipertimbangkan. Pada tahap pertama, model pengujian adalah dikembangkan untuk mengidentifikasi posisi jeda di dalam sebuah ucapan. Pada tahap kedua, jeda diklasifikasikan menjadi tiga tipe berbeda menggunakan fitur suku kata. Di final tahap, prediktor regresi dimodelkan untuk memprediksi durasi jeda ini berdasarkan jenisnya. Model CART dievaluasi pada setiap tahap dengan menggunakan langkah-langkah obyektif. Tes mendengarkan subyektif juga dilakukan setelah dimasukkan metode yang diusulkan dalam TTS Hindi dasar. Hasil dari evaluasi persepsi menunjukkan bahwa metode yang diusulkan efektif dalam memaksakan jeda dalam pidato yang disintesis ucapan.

Tabel 2.1 Tabel Perbandingan Penelitian

No	Penulis/Judul	Metode	Hasil
1.	<ul style="list-style-type: none"> • Krishnendu Ghosh and K. Sreenivasa Rao (2012) • <i>Data-Driven Phrase Break Prediction for Bengali Text-to-Speech System</i> 	<ul style="list-style-type: none"> • Membandingkan pendekatan <i>Feed-Forward Neural Network</i> (FFNN) dan <i>Classification and Regression Tree</i> (CART) model. • Korpus yang digunakan adalah korpus bahasa Bengali India sebanyak 6.210 kalimat. 	<ul style="list-style-type: none"> • Hasil dari penelitian tersebut dengan menggunakan CART nilai <i>precision</i> 80.66%, <i>recall</i> 75.05%, dan <i>f-measure</i> 77.76%. • Menggunakan FFNN nilai <i>precision</i> 82,66%, <i>recall</i> 79.87%, <i>f-measure</i> 81.06%.

No	Penulis/Judul	Metode	Hasil
2.	<ul style="list-style-type: none"> Aditya Teguh Nugraha (2014) Prediksi Jeda Dalam Ucapan Kalimat Bahasa Indonesia Dengan <i>Hidden Markov Model</i> 	<ul style="list-style-type: none"> Menggunakan <i>Hidden Markov Model</i> Melakukan penelitian untuk prediksi jeda kalimat Bahasa Indonesia dengan <i>Hidden Markov Model</i> (HMM). Terdapat dua proses utama dalam aplikasi ini yaitu train untuk melatih korpus, dan prediksi untuk memprediksi jeda. 	<ul style="list-style-type: none"> Hasil dari penelitian tersebut adalah kalimat yang telah diprediksi kejadian jedanya dengan nilai <i>precision</i> sebesar 0.364, <i>recall</i> sebesar 0.132, dan <i>f-score</i> sebesar 0.194.
3.	<ul style="list-style-type: none"> Muhammad Iqbal Kamaludin (2017) Prediksi Jeda Dalam Ucapan Bahasa Melayu Pontianak Dengan Menggunakan Metode <i>Shallow Parsing</i> 	<ul style="list-style-type: none"> Melakukan penelitian untuk prediksi jeda pada ucapan Bahasa Melayu Pontianak dengan menggunakan PoS <i>tagging</i> Bahasa Melayu Pontianak. Menggunakan <i>shallow parsing</i> berdasarkan <i>rule grammar</i> Bahasa Melayu Pontianak. 	<ul style="list-style-type: none"> Dari 168 kalimat yang merupakan gabungan dari kalimat tunggal dan kalimat majemuk menghasilkan nilai kecocokan dengan jeda penutru sebesar 40.4% atau 68 kalimat. Hasil dari penelitian tersebut menghasilkan nilai recall 0.78 untuk kalimat tunggal 0.67 untuk kalimat majemuk yang menunjukkan bahwa sebagian besar frasa jeda yang relevan tidak terbentuk oleh <i>shallow parsing</i>. Nilai <i>precision</i> 0.74 untuk kalimat tunggal

No	Penulis/Judul	Metode	Hasil
.			dan 0.57 untuk kalimat majemuk

2.2 *Natural Language Processing*

Natural Language Processing (NLP) adalah sebuah bagian dari penelitian dan aplikasi yang mengkaji bagaimana komputer dapat digunakan untuk mengerti dan memanipulasi bahasa alami yang berupa teks atau ucapan untuk hal –hal yang berguna (Kamiludin, 2017). Manipulasi teks telah dikenal sebagai sebuah bidang penelitian yang penting dalam NLP. Sebuah sistem NLP yang mengolah teks dimulai dengan analisis morfologi. Teks dikonversi, dalam kueri atau dokumen, untuk mendapatkan varian morfologi kata-kata yang terlibat. Pengolahan leksikal dan sintaktis melibatkan pemanfaatan kamus untuk menentukan karakteristik dari kata-kata, pengenalan *part of speech*, menentukan kata-kata dan frasa, serta untuk penguraian kalimat (Chowdhury, 2003).

2.3 **Informasi Parafrasa**

Parafrasa yaitu ungkapan kembali maksud atau isi tulisan dengan menggunakan kata-kata sendiri (Sabarti, 1996). Kata lain dari parafrasa yaitu mengungkapkan kembali berbagai informasi yang dibaca atau yang didengar dengan cara dan bentuk yang berbeda. Selain itu, parafrasa juga digunakan untuk menguraikan suatu teks dalam bentuk atau susunan kata yang lain agar makna yang tersembunyi dalam teks tersebut dapat dijelaskan.

Informasi parafrasa diambil berdasarkan kata kunci dan tipe PoS. Kata kunci adalah kata yang akan ditetapkan dan digunakan dalam prediksi jeda dengan menggunakan tipe PoS. Kunci digunakan untuk memasukan jeda antara kalimat yang meningkatkan kejelasan dan kealamiannya. Kata kunci adalah kata yang telah ditetapkan yang akan digunakan untuk pemberian jeda. Pemberian jeda pada kata kunci dapat terjadi sebelum atau sesudah kata yang berada pada korpus, kemudian kata tersebut yang akan dimasukkan dalam tabel kata kunci.

Aturan kata kunci didapatkan dari hasil penandaan jeda. Penandaan jeda didapatkan dari penandaan *wavesurfer*. Kriteria analisis untuk pembuatan aturan kata kunci sebagai berikut:

- a. Analisis dilakukan dengan menggunakan 500 kalimat Bahasa Melayu Pontianak yang sudah ditandai indeks jeda sesuai dengan kejadian jeda dari penutur.
- b. Analisis dilihat dari kejadian jeda di tiap frasa dalam kalimat yang sudah ditandai indeks jeda sesuai dengan kejadian jeda dari penutur.
- c. Analisis dilihat berdasarkan kata pertama yang terdapat kejadian jeda.
- d. Analisis dilakukan dengan menghitung seberapa banyak jumlah kemunculan kata setiap jeda yaitu dengan melihat jumlah kemunculan kata setelah jeda, jumlah kemunculan kata sebelum jeda, dan jumlah kemunculan kata tidak ada jeda.
- e. Analisis tidak menggunakan minimal berapa banyak kemunculan jeda yang terjadi.
- f. Hasil analisis ditentukan dari jumlah terbanyak kemunculan jeda.

2.4 *Part Of Speech Tagging*

Part of Speech (PoS) Tagging , yang juga disebut sebagai pelabelan kelas kata, adalah suatu proses yang memberikann label kelas kata secara otomatis pada suatu kata dalam kalimat (Jurafski, 2000). Kelas kata terdiri dari kata sifat, kata benda, kata kerja, kata keterangan, kata depan, kata kepemilikan, kata penghubung, dll. Suatu kata bisa diklasifikasikan ke dalam kelas kata yang berbeda (Iqbal Kamiluddin, 2017).

Berbagai set PoS telah digunakan dalam penelitian-penelitian bidang pemrosesan bahasa alami (PBA), diantaranya lewat PAN Localization Project, khusus untuk bahasa indonesia telah dikembangkan PoS khusus untuk terjemahan ke bahasa Inggris pada tahun 2009 (Adriani, 2009). Kelas kata tersebut dibangun berdasarkan Penn Treebank PoS *tag sets* (Marcus dkk.,1993) terdiri dari 29 PoS tag. PoS utama bahasa indonesia adalah kata kerja (*verb*), kata sifat (*adjective*), kata tugas (*function words*), kata keterangan (*adverb*), dan kata benda (*noun*), berdasarkan 5 (lima) PoS utama tersebut dan observasi data, (Pisceldo dkk,2009) mendefinisikan 37 tag hasil dari modifikasi tag set yang di hasilkan oleh Adriani, (2009) dan Pisceldo dkk. (2009). Herry sujaini (2014) dalam penelitiannya menggunakan 42 *tag* untuk Bahasa Indonesia Penelitian ini menggunakan tipe set PoS dari penelitian sebelumnya, yaitu set PoS dari Alfian Farizki Wicaksono

(HMM Based Part-of-Speech Tagger for Bahasa Indonesia:2010)

Table 3. Indonesian PoS Set Recommended by Computational Based

No	Tag	Description	Word Examples
1	OP	Opening parenthesis	{
2	CP	Closing parenthesis	}
3	GM	Slash	/
4	:	Semicolon	:
5	:	Colon	:
6	"	Quotation	"
7	.	Sentence terminator	.
8	,	Comma	,
9	-	Dash	-
10	...	Ellipsis	...
11	JJ1	Adjectives 1	panjang, kuat, indah, besar
12	JJ2	Adjectives 2	genap, buntu, negatif
13	RB	Adverbs	sekedar, hampir, tidak
14	RB1	Adverbs 1	sangat, amat, cukup, paling
15	NN	Common Noun	mobil, air, negara
16	NNP	Proper nouns	tvri, jokowi, persib
17	NNG	Genitive nouns	bukunya, hatinya
18	VBI	Intransitive Verb	duduk, pergi, makan
19	VBI1	Intransitive Verb 1	dibuat, diambil
20	VBI2	Intransitive Verb 2	mengecil, menguat
21	VBT	Transitive Verb	membeli, memukul
22	IN	Preposition	di, ke, dari
23	MD	Modal	akan, harus
24	CC	Coor - conjunction	dan, atau, ketika, jika
25	DT	Determiner	ini, itu
26	UH	Interjections	wah, aduh, oi
27	CDO	Ordinal numerals	pertama, kedua
28	CDC	Collective numerals	berdua, bertiga
29	CDP	Primary numerals	1, 2, 3
30	CDP1	Primary numerals 1	satu, dua
31	CDP2	Primary numerals 2	puluh, ribu, juta
32	CDP3	Primary numerals 3	1990, 2001, 2013
33	CDI	Irregular numerals	beberapa
34	PRP	Personal pronoun	saya, kamu
35	WP	WH-pronouns	apa, siapa
36	PRN	Number pronouns	kedua-duanya
37	PRL+	Locative Proper nouns/pronouns	sini, situ, Jakarta, Bali
38	SYM	Symbols	@#\$%^&
39	RP	Particles	pun, kah
40	FW	Foreign words	foreign, Word
41	ART	Articles	sang, si, para
42	COP	Copula	adalah, bukan, merupakan

Iqbal Kamiluddin (2017) dalam penelitiannya mengenai prediksi jeda menggunakan *shallow parsing* menggunakan 35 tag PoS dari Alfian Farizki Wicaksono (2010) dan dilakukan secara

Gambar 2. 1 Set PoS Herry Sujaini (2014)

manual sebanyak 168 kalimat Bahasa

Tabel tag PoS Bahasa Melayu Pontianak seperti pada tabel 2.2 berikut:

Tabel 2.2 PoS Tag Bahasa Melayu Pontianak M.Iqbal Kamiluddin (2017)

No	PoS	Deskripsi	Contoh Kata
1	OP	<i>Parenthesis</i>	{
2	CP	<i>Close Parenthesis</i>	}
3	GM	<i>Slash</i>	/

No	PoS	Deskripsi	Contoh Kata
4	;	<i>Semicolon</i>	;
5	:	<i>Colon</i>	:
6	"	<i>Quotation</i>	" "
7	.	<i>Sentence Terminator</i>	. ! ?
8	,	<i>Comma</i>	,
9	-	<i>Dash</i>	-
10	...	<i>Ellipsis</i>	...
	JJ	<i>Adjective</i>	Kaya, Manis
11	RB	<i>Adverb</i>	Nanti, Sementara
	NN	<i>Common Noun</i>	Mobil, Sepatu, Roda
12	NNP	<i>Proper Noun</i>	Bekasi, Indonesia
	NNG	<i>Genitive Noun</i>	Bukunya
13	VBI	<i>Intransitive Verb</i>	Pergi
14	VBT	<i>Transitive Verb</i>	Membeli
15	IN	<i>Preposition</i>	Di, Ke, Dari
16	MD	<i>Modal</i>	Bisa
17	CC	<i>Coor-Conjunction</i>	Dan, Atau, Tetapi
18	SC	<i>Subor-Conjunction</i>	Jika, Ketika
19			
20			
21			
22	DT	<i>Determiner</i>	Para, Ini, Itu
	UH	<i>Interjection</i>	Wah, Aduh, Oi
23	CDO	<i>Ordinal Numerals</i>	Pertama, Kedua
	CDC	<i>Collective Numerals</i>	Bertiga
24	CDP	<i>Primary Numerals</i>	Satu, Dua
	CDI	<i>Irregular Numerals</i>	Beberapa
25	PRP	<i>Personal Pronouns</i>	Saya, Kamu
	WP	<i>WH-Pronouns</i>	Apa, Siapa, Dimana
26	PRN	<i>Number Pronouns</i>	Kedua-duanya, Satu-
	PRL	<i>Locative Pronouns</i>	satunya
27	NEG	<i>Negation</i>	Sini , Situ, Sana
	SYM	<i>Symbols</i>	Bukan, Tidak
28	RP	<i>Particles</i>	@#\$%^&
	FW	<i>Foreign</i>	Pun, Kah
29			Words
30			
31			
32			
33			
34			
35			

2.5 WaveSurfer

WaveSurfer adalah sebuah aplikasi *Open Source* untuk visualisasi dan manipulasi suara. Aplikasi ini telah dirancang untuk memenuhi kebutuhan pemula dan pengguna tingkat lanjut. WaveSurfer memiliki antarmuka yang sederhana dan menyediakan fungsi dalam cara yang intuitif dan yang dapat disesuaikan dengan berbagai tugas. WaveSurfer dapat digunakan sebagai alat yang berdiri sendiri untuk berbagai macam tugas penelitian ucapan dan pendidikan. Jenis pemakaian yang umum digunakan adalah analisis ucapan/suara dan anotasi/transkripsi suara. (Nugraha Teguh, 2014)

2.6 Korpus

Konsep korpus seperti itu merupakan pengertian korpus secara tradisional. Saat ini konsepsi korpus telah berkembang dan bermakna luas. Sekarang, korpus juga dapat dikumpulkan secara elektronik. Ada tiga aspek yang menjadi pertimbangan dalam memahami konsep (Baker, 1995). Pertama, korpus utamanya merupakan kumpulan teks yang dihasilkan secara elektronik dan dapat dianalisis secara otomatis atau semi otomatis. Kedua, korpus tidak hanya berisi kumpulan teks bentuk tulis, tetapi juga mencakup ujaran. Ketiga, korpus mungkin juga mencakup sejumlah besar teks yang berasal dari beragam sumber, misalnya dari beragam penulis dan penutur dan dalam berbagai topik (Baker, 1995).

Korpus didefinisikan sebagai koleksi atau sekumpulan contoh teks tulis atau lisan dalam bentuk yang dapat dibaca dengan menggunakan seperangkat mesin dan dapat diberi catatan berupa berbagai bentuk informasi linguistik (McEnery dkk, 2006).

McEnery (2006) merangkum kriteria korpus, yang telah menjadi kesepakatan banyak ahli, yakni:

1. dapat dibaca dengan menggunakan seperangkat mesin,
2. berupa teks otentik,
3. digunakan sebagai sampel,
4. mewakili bahasa atau variasi bahasa tertentu.

Korpus atau data yang dikumpulkan berupa text kalimat Bahasa Melayu Pontianak yang di dapatkan dari buku Pay Jarot Sujarwo dengan judul Sepok Satu dan Sepok Tige.

Contoh korpus yang digunakan:

1. Ha, ngaku gak kau aher-e, tak usahlah kau nak nyangkal agek.
2. Manelah negare kau tuh nak maju kalok tebiat pemerintah-e tak tentu rudu macam itu.
3. Dahlah kite tunggu jak beberape ari agik.
4. Aoklah, yang sabar jak ye, aku nak op lok.

2.7 Bahasa Melayu Pontianak

Etnis Melayu adalah etnis orang-orang Austronesia yang sebagian besar menempati daerah semenanjung Malaka, Sumatera, daerah selatan Thailand, Pantai Selatan Burma, Singapura, Brunei, Kalimantan Barat, Serawak, dan pulau-pulau kecil yang berada di sekitar lokasi-lokasi tersebut. Lokasi-lokasi itu, sekarang adalah bagian dari negara-negara, seperti Malaysia, Indonesia, Singapura, Brunei, Burma, dan Thailand. Terdapat beberapa perbedaan dalam *genetis*, *linguistik*, budaya, dan sosial di beberapa subetnis Melayu, yang disebabkan terjadinya imigrasi dan asimilasi dari berbagai etnis ratusan yang lalu (Collins, 2005).

Bahasa Melayu Pontianak merupakan dialek bahasa Melayu yang dituturkan oleh masyarakat Kota Pontianak, Kabupaten Kubu Raya dan Kabupaten Mempawah serta memiliki kesamaan dengan bahasa Melayu Semenanjung Malaysia (Johor-Riau). Bahasa Melayu Pontianak dipengaruhi oleh bahasa Dayak dari rumpun Klementan juga memiliki kesamaan beberapa kosakata dengan bahasa Melayu yang dituturkan di wilayah Kabupaten Sambas, Kabupaten Kayong Utara, dan Kabupaten Ketapang (Putra Etsa, 2018). Dalam bahasa Melayu Pontianak tidak mengenal tingkatan berbahasa seperti halus, sebaya atau kasar seperti pada bahasa Jawa, karena kasar dan halusnya pengucapan bahasa Melayu Pontianak tergantung pada intonasi dan penekanan nada. Penggunaan bahasa Melayu Pontianak juga sering disingkat dari kata aslinya pada pembicaraan sehari-hari. Untuk mengetahui jumlah asli penutur bahasa ini memang sulit didapat, karena tidak ada data yang pasti mengenai jumlah penutur. Adapun contoh Bahasa Melayu Pontianak seperti berikut:

“Bia-pon kate orag duak bulan lebéh tuh waktu yang lamak untok diam di Belande, tapi menurot aku tetap jak tak cukup. Suke rase-e ati ni. Néngok gedong-gedong tue. Néngok parét-parét canték mempesone. Ngerasekan hawe séjuk. Néngok pokok-pokok tadak bedaon. Tadak cume itu, di Belande pon aku belajar

banyak tegal disiplin. Péndek-e, aku betah diam di Belande. Ape agék aku jumpe banyak indomie, mangkén nyaman lah lior aku dibuatkan-e. Makan indomie pakau kéju, suah ndak kitak? Aku sih suah. Macam mane rase-e? Kate budak Kampong Arang, pandai jak perot ngodak-e. Pakal ba aku ni ade keje, akibat-e tadak beduét. Jadi tadak bise lamak-lamak betanggam di Belande. Kok lamak-lamak, bise medam palak”.

2.8 Pengujian Sistem

2.8.1 *Precision & Recall*

Precision adalah rasio jumlah dokumen relevan yang ditemukan dengan jumlah total dokumen yang ditemukan oleh sistem. *Recall* adalah rasio jumlah dokumen relevan yang ditemukan kembali dengan total jumlah dokumen dalam kumpulan dokumen yang dianggap relevan (Rozi Fahrur, 2012). Persamaan untuk *precision* ditunjukkan pada persamaan dan *recall* pada persamaan

Tabel 2.3 Variabel Untuk Perhitungan *Precision* dan *Recall*

	Relevan	Tak Relevan
Ditemukan	<i>True positives (tp)</i>	<i>False positives (fp)</i>
Tidak ditemukan	<i>False negatives (fn)</i>	<i>True negatives (tn)</i>

$$\text{Precision} = \frac{\text{item yang relevan yang ditemukan}}{\text{item ditemukan}} = P(\text{relevan}|\text{ditemukan})$$

$$\text{Recall} = \frac{\text{item yang relevan yang ditemukan}}{\text{semua item yang relevan}} = R(\text{ditemukan}|\text{relevan})$$

Dengan memperhatikan Tabel 2.5, dapat disederhanakan menjadi :

$$\text{Precision} = \frac{a}{a+c}$$

$$\text{Recall} = \frac{a}{a+b}$$

Pengujian *recall* dan *precision* digunakan untuk mengetahui nilai akurasi keberhasilan penelitian ini yang dibuat dalam menghasilkan potongan kalimat yang terbagi dalam kata kunci. Pengujian ini juga digunakan sebagai nilai acuan untuk pengujian kata kunci terhadap frasa jeda yang didapat dari penutur apakah bisa menghasilkan frasa jeda yang baik sesuai penutur Bahasa Melayu Pontianak.

2.8.2 *F-Measure*

F-Measure (nilai F) merupakan sebuah nilai dari keakuratan sebuah tes. Nilai F menggunakan *precision* & *recall* dari tes untuk menghitung nilainya, dengan *precision* yang menyatakan jumlah hasil benar dibandingkan dengan jumlah hasil yang ditemubalikkan dan *recall* yang menyatakan jumlah hasil benar dibandingkan dengan jumlah hasil yang harus ditemubalikkan. Nilai F dapat diartikan sebagai sebuah hasil rata-rata dari *precision* & *recall*, dimana sebuah nilai F mendapat hasil 1 sebagai hasil terbaik dan 0 sebagai nilai terburuk (Kamliluddin, 2017). Rumus nilai F adalah sebagai berikut:

$$F\text{-measure} = 2 \cdot \frac{\textit{precision} \cdot \textit{recall}}{\textit{precision} + \textit{recall}}$$

BAB III

METODOLOGI PENELITIAN

3.1 Bahan Penelitian

Bahan Penelitian berupa korpus kalimat Bahasa Melayu Pontianak yang didapat dari buku Pay Jarot Sujarwo dengan judul *Sepok Tige*. Kalimat teks sudah ditandai jedanya oleh penutur Bahasa Melayu Pontianak.

3.2 Perangkat Penelitian

Perangkat penelitian yang digunakan berupa alat penelitian, perangkat keras, dan perangkat lunak. Adapun penjabaran perangkat penelitian yang dipergunakan adalah sebagai berikut:

3.2.1 Perangkat Keras

Perangkat keras yang digunakan dalam penelitian adalah sebuah laptop Asus X45U dengan spesifikasi AMD E-450 APU with Radeon HD Graphics ~1.6GH dengan Harddisk 500 GB dan kapasitas RAM 4 GB.

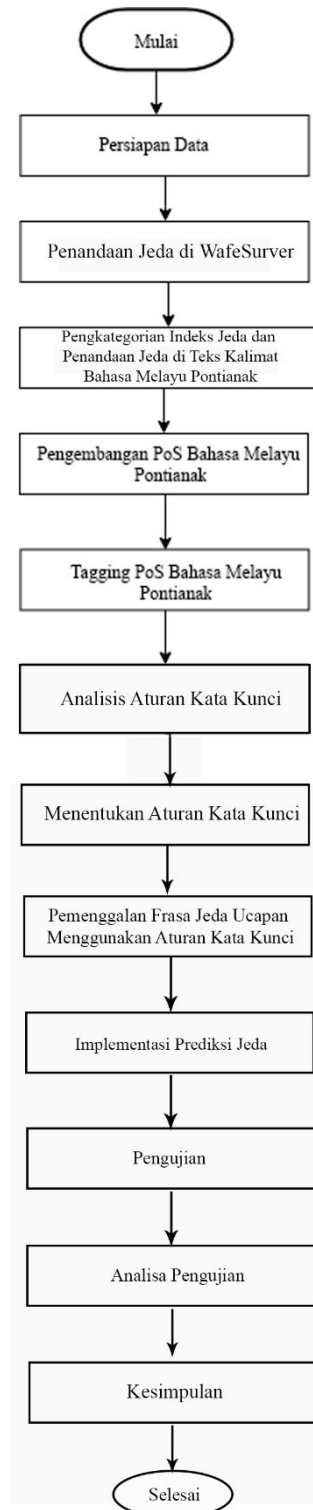
3.2.2 Perangkat Lunak

Perangkat lunak yang digunakan dalam penelitian ini adalah:

- a. Sistem Operasi Windows 10 Pro, 64 bit.
- b. *WafeSurver* 1.8.8p4 sebagai aplikasi untuk mendengarkan frasa ucapan dan tagging jeda.
- c. *Python* 3.6.5 sebagai aplikasi sistem yang akan dibuat.
- d. Microsoft Excel 2013 sebagai pengelola data.
- e. Sublime Text 3 v3126 sebagai aplikasi *text editor*.

3.3 Metode Penelitian

Metode penelitian yang akan dilakukan akan dijelaskan pada diagram alir penelitian pada gambar 3.1 berikut :



Gambar 3.1 Diagram Alir Penelitian

Seperti diagram alir penelitian yang terlihat pada gambar 3.1, terdapat beberapa tahapan yang dilakukan pada penelitian, diantaranya dimulai dari persiapan data berupa *file* teks dan *file* rekaman suara penutur bahasa Melayu

Pontianak yang berjumlah 500 kalimat, penandaan kejadian jeda di *Wavesurfer*, pengkategorian indeks jeda dan penandaan jeda di korpus, pengembangan PoS bahasa Melayu Pontianak, melakukan PoS Tagging bahasa Melayu Pontianak, analisis aturan Kata Kunci dan menentukan aturan kata kunci, melakukan pengujian, analisa hasil pengujian dan yang terakhir melakukan penarikan kesimpulan.

3.3.1 Persiapan Data

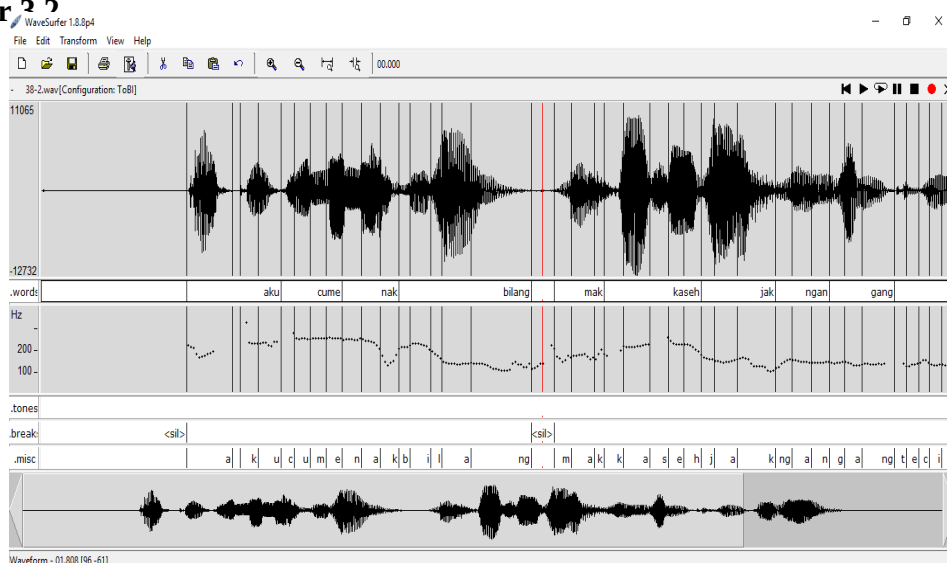
Semua data yang diperlukan berupa korpus bahasa Melayu Pontianak yang diambil dari buku Sepok 1 karya Pay Jarrot Sujarwo. Tidak semua teks dari buku tersebut digunakan. Beberapa hal yang perlu disaring dalam teks cerita tersebut diantaranya tidak menyertakan kalimat yang terdapat kata atau istilah selain bahasa Melayu Pontianak. Korpus Bahasa Melayu Pontianak yang diambil sebanyak 500 kalimat.

3.3.2 Penandaan Kejadian Jeda di *Wavesurfer*

File suara yang digunakan merupakan rekaman dari teks yang ada pada buku Sepok Satu karangan Pay Jarot Sujarwo dan dilakukan oleh penutur ahli. *File* suara ini berformat .wav dengan resolusi 16 bit. *File* suara tersebut kemudian dilakukan penandaan jeda kata per kata meggunakan aplikasi WaveSurfer dengan mendengarkan jeda yang terjadi pada ucapan penutur.

Gambar 2.2

Proses



penandaan jeda/SIL di Wavesurfer

Dapat dilihat pada gambar 3.2 jeda yang ditandai dengan “|<sil>|” pada bagian “break”. Tanda “|” sebelum <sil> menandakan jeda awal, sedangkan tanda “|” setelah <sil> menandakan jeda akhir. Hasil dari “break” diolah untuk dijadikan durasi jeda pada indeks jeda.

3.3.3 Pengkategorian Indeks Jeda dan Penandaan Jeda di Teks Kalimat Bahasa Melayu Pontianak

Setelah semua file suara ditandai, selanjutnya setiap kejadian jeda dikategorikan jeda menggunakan indeks jeda. Penelitian ini menggunakan 5 kategori indeks jeda, yaitu “0” untuk indeks yang menyatakan tidak ada jeda, “1” untuk indeks yang menyatakan terjadinya jeda singkat, “2” untuk indeks yang menyatakan terjadinya jeda panjang, “,” untuk indeks yang menyatakan tanda baca koma, dan “.” untuk indeks yang menyatakan akhir kalimat.

Indeks Jeda	Keterangan
0	Tidak ada jeda
1	Jeda yang terjadi tipis
2	Jeda yang terjadi tebal
,	Tanda baca koma (,)
.	Akhir kalimat

Tabel 3.1 Kategori Indeks Jeda

Indeks jeda pada tabel 3.1 , selanjutnya akan ditentukan lama durasi jeda yang terjadi. Perhitungan durasi jeda didapatkan dengan cara menghitung nilai *mean* atau rata-rata, modus, nilai minimum, dan nilai maksimum dari kejadian jeda di file *.breaks yang telah diurutkan dari nilai yang terkecil hingga terbesar dengan menggunakan bantuan *spreadsheet* (Nugraha, 2014).

Jeda awal	Jeda Akhir	Durasi Jeda	Jumlah Kemunculan
...

Tabel 3.2 Tabel Proses Perhitungan Durasi Jeda

Durasi jeda didapat dengan mengurangkan nilai jeda akhir dan jeda awal. Nilai di kolom durasi jeda dihitung untuk mendapatkan nilai MIN, MAX, Median

dan Mean atau rata-rata. Sedangkan nilai di kolom jumlah kemunculan digunakan sebagai nilai modus. Nilai tersebut digunakan sebagai nilai durasi untuk indeks jeda sehingga hasilnya dapat dilihat pada tabel 3.3.

Tabel 3.3 Kategori Indeks Jeda dan Durasinya

Indeks Jeda	Keterangan	Durasi Jeda
0	Tidak ada jeda	0 - < nilai modus
1	Jeda yang terjadi tipis	nilai modus - <= nilai <i>mean</i>
2	Jeda yang terjadi tebal	> nilai <i>mean</i>
,	Tanda baca koma (,)	,
.	Akhir kalimat	.

Indeks jeda 0 memiliki durasi jeda dari nol detik sampai kurang dari nilai modus. Indeks jeda 1 atau jeda singkat memiliki durasi jeda dari nilai modus sampai kurang dari sama dengan nilai mean. Sedangkan untuk indeks jeda 2 atau jeda panjang memiliki durasi jeda lebih dari nilai *mean*. Kategori indeks jeda singkat dan panjang beserta durasi yang telah didapat kemudian ditandai kembali di korpus teks bahasa Melayu Pontianak sesuai lama jeda yang telah ditandai di *wavesurfer*.

3.3.4 Pengembangan PoS bahasa Melayu Pontianak

Pada tahap ini digunakan set PoS yang baru yang dibentuk berdasarkan Set PoS Bahasa Melayu Pontianak. Pengelompokkan kelas kata dalam Bahasa Melayu dilakukan secara manual dengan merujuk pada referensi set PoS pada penelitian M.Iqbal Kamiludin (2017), dan beberapa referensi set PoS yang ada yakni dari Set Pos Setyaningsih (2017) dan Set PoS Adriani (2009). Set PoS baru dibuat dengan tujuan untuk memperjelas penandaan tipe kata dalam kalimat dan meningkatkan akurasi prediksi jeda.

Pengembangan PoS bahasa Melayu Pontianak dilakukan dengan cara menganalisa tabel PoS referensi yang ada dan mencocokkan label PoS yang sesuai dengan kelas kata yang digunakan pada bahasa Melayu Pontianak. Penambahan set PoS disesuaikan dengan fungsi kata dalam kalimat bahasa Melayu Pontianak. Misalkan, pada penelitian M.Iqbal Kamiludin, kata “si” belum memiliki label PoS

yang tepat yaitu NN. Pada penelitian Setyaningsih (2017) , terdapat label PoS AR atau artikel untuk panggilan seperti “si”, “sang”, atau “Yang”, sehingga PoS AR diambil dan ditambahkan ke tabel PoS yang baru. Untuk kata keterangan atau *adverb*, pada penelitian sebelumnya hanya digunakan PoS RB menandai kata jarang, sementara, dan agak. Ketiga kata ini sebenarnya memiliki keterangan yang berbeda dan bisa dikelompokkan menjadi PoS baru untuk memperjelas tipe kelas kata. Ketiga kata dapat dikembangkan menjadi PoS RB atau kata keterangan umum untuk kata “sementara”. Kata “jarang” menandai keterangan waktu seseorang melakukan kegiatan sehingga dibuat PoS FRB atau kata keterangan untuk kata yang mendeskripsikan seberapa sering suatu kejadian terjadi atau dilakukan. Kata “agak” menunjukkan keterangan perkiraan berapa lama suatu kegiatan sehingga dibuat PoS DRB atau kata keterangan tingkatan. Hasil dari pengembangan PoS adalah tabel set PoS baru yang digunakan untuk menandai kata di korpus teks bahasa Melayu Pontianak.

3.3.5 Tagging PoS bahasa Melayu Pontianak

Pada dasarnya, IPOSTagger menggunakan korpus Bahasa Indonesia. Maka dari itu, korpus yang ada harus diganti dengan korpus Bahasa Melayu Pontianak. Teks kalimat untuk korpus latih IPOSTagger yang berasal dari buku Sepok Satu diberikan *tag* PoS sesuai dengan kelas katanya secara manual sesuai set PoS yang telah dikembangkan. Penandaan PoS berdasarkan makna dari kata tersebut ketika diucapkan dan dari Kamus Besar Bahasa Indonesia (KBBI). Korpus latih tersebut berisikan “kata/PoS”.

Misalkan pada korpus terdapat kalimat :

“Eh , jangan banyak umong kau .”

Setiap kata akan dilihat dan dikelompokkan berdasarkan label PoS. Misalkan kata “Eh” menunjukkan panggilan pada kalimat dengan kata seru sehingga ditandai dengan PoS “UH” untuk kata seru. Kata “jangan” menunjukkan kata negatif sehingga ditandai dengan PoS “NEG” untuk kata negatif. Kata “banyak” menunjukkan kata sifat sehingga ditandai dengan PoS “JJ” untuk kata sifat. Kata “ngumong” menunjukkan kegiatan berbicara sehingga ditandai dengan PoS “VBI” untuk kata kerja intrasitif yang memerlukan objek dibelakang kata kerja. Kata “kau” merupakan objek yang berupa kata ganti orang sehingga

ditandai dengan PoS “PRP” yang dikhususkan untuk kata ganti orang seperti “kau”, “aku” , “kamek”. Hasil penandaan PoS menjadi kalimat yang berisi “kata/PoS” seperti berikut:

“Eh/UH ,/, jangan/NEG banyak/JJ umong/VBI kau/PRP ./.”

3.3.6 Analisis dan Menentukan Aturan Kata Kunci

Setelah melakukan Tagging PoS pada kalimat Bahasa Melayu Pontianak, selanjutnya kalimat Bahasa Melayu Pontianak dianalisis untuk mendapatkan *aturan* kata kunci yang diperlukan dalam melakukan proses prediksi jeda.

Aturan kata kunci dilakukan untuk mendapatkan hasil dari penandaan jeda. Penandaan jeda didapatkan dari penandaan *wavesurfer*. Kriteria analisis untuk pembuatan aturan kata kunci sebagai berikut:

- a. Analisis dilakukan dengan menggunakan 500 kalimat Bahasa Melayu Pontianak yang sudah ditandai indeks jeda sesuai dengan kejadian jeda dari penutur.
- b. Analisis dilihat dari kejadian jeda di tiap frasa dalam kalimat yang sudah ditandai indeks jeda sesuai dengan kejadian jeda dari penutur.
- c. Analisis dilihat berdasarkan kata pertama yang ditandai jeda.
- d. Analisis dilakukan dengan menghitung seberapa banyak jumlah kemunculan kata setiap jeda yaitu dengan melihat jumlah kemunculan kata setelah jeda, jumlah kemunculan kata sebelum jeda, dan jumlah kemunculan kata tidak ada jeda.
- e. Analisis tidak menggunakan minimal berapa banyak kemunculan jeda yang terjadi.
- f. Hasil analisis ditentukan dari jumlah terbanyak kemunculan jeda .

Contoh kalimat untuk analisis kata kunci sebagai berikut:

1. Aku lupa-lupa ingat/2 , tapi/1 kalok tak salah sih/2 waktu itu aku tidok belior .
2. Lalulah abang itu/1 ku suroh balek.
3. Sari itu/2 aku dah janji ngan kitak kan/2 kalok nak nceritekan soal sepeda ini .

4. Tapi rupe-e/2 itu/1 tak bise jadi solusi .

Tabel 3.4 Analisis Kemunculan Aturan Kata Kunci

No.	Kata Kunci	Jumlah Kemunculan Kata Setelah Jeda	Jumlah Kemuculan Kata Sebelum Jeda	Jumlah Kemuculan Kata Tidak Ada Jeda	PoS
1.	Tapi		1		CON
2.	Sih		1		RP
3.	Itu	1	3	1	DT
4.	Kan		1		RP
5.	Rupe-e		1		RB

Kemunculan aturan kata kunci dilakukan dengan cara menghitung setiap kejadian jeda pada setiap kata. Kata “tapi” pada kalimat pertama terjadi setelah jeda, maka kata “tapi” dimasukkan pada kolom kemuculan setelah jeda. Kata “itu” pada kalimat pertama tidak terdapat jeda maka kata “itu” dimasukkan pada kolom kemunculan tidak ada jeda, pada kalimat kedua kata “itu” terjadi setelah jeda maka kata “itu” juga dimasukkan kedalam kolom kemuculan setelah jeda, pada kalimat ketiga kata “itu” terjadi setelah jeda maka kata “itu” juga dimasukkan kedalam kolom kemuculan setelah jeda, dan pada kalimat keempat kata ”itu” terjadi sebelum dan setelah jeda maka kata “itu” dimasukkan kedalam kolom kemunculan sebelum dan setelah jeda. Analisis dilakukan menggunakan 500 kalimat Bahasa Melayu Pontianak dan hasil yang akhirnya dipilih untuk aturan kata kunci adalah kejadian jeda yang paling banyak muncul, yang membedakan dengan menggunakan PoS yaitu setiap katanya dibagi menjadi kelas kata karena ada kata sama tapi tipe PoS nya berbeda. Penggunaan set PoS bisa dilihat pada tabel 2.4 Bab 2 tentang set PoS Bahasa Melayu Pontianak M.Iqbal Kamiludin.

Tabel 3.5 Kemunculan Jeda

No.	Kata Kunci	Jumlah Kemuculan Jeda 1	Jumlah Kemunculan Jeda 2
1.	Tapi	1	
2.	Sih		1
3.	Itu	2	2

4.	Kan	1	
5.	Rupe-e	1	

Kemunculan jeda terbagi menjadi 2 yaitu jeda 1 (singkat) dan 2 (panjang). Pada kalimat pertama kemunculan jeda bisa dilihat pada kata “tapi”, kata “tapi” terjadi pada jeda 1 (singkat) maka kata “tapi” dimasukkan pada kolom kemunculan jeda 1, pada kalimat yang sama ada kata “sih”, kata “sih” terjadi pada jeda 2 (panjang) maka kata “sih” dimasukkan pada kolom kemunculan jeda 2. Pada kalimat kedua dan keempat ada kata “itu”, kata “itu” terjadi pada jeda 1 (singkat) maka kata “itu” dimasukkan pada kolom kemunculan jeda 1, sedangkan pada kalimat ketiga kata “itu” terjadi pada jeda 2 (panjang) maka kata “itu” dimasukkan pada kolom kemunculan jeda 2. Hasil yang dipilih adalah kejadian jeda yang paling banyak muncul.

3.3.7 Pemenggalan Frasa Jeda Ucapan Menggunakan Aturan Kata Kunci

Tahap yang selanjutnya dilakukan adalah melakukan pemenggalan frasa jeda menggunakan *aturan* kata kunci yang telah dibuat. Cara pemenggalan frasa ucapan jeda, yakni dengan memproses kalimat inputan menjadi penggalan frasa sesuai *aturan* kata kunci. Diagram alir proses pemenggalan frasa ucapan jeda dengan aturan kata kunci dapat dilihat pada gambar 3.3.



Gambar 3.3 Diagram Alir Pemenggalan Jeda dengan Aturan Kata Kunci

Tahapan proses pemenggalan jeda dengan aturan kata kunci pada gambar 3.3 terdiri dari input kalimat, pengecekan aturan kata kunci dan pemenggalan frasa sesuai aturan kata kunci dan kemunculan jeda serta *output* yang berupa penggalan

frasa jeda ucapan.

3.3.7.1 Input

Proses yang pertama kali dilakukan dalam pemenggalan frasa jeda dengan aturan kata kunci adalah membaca input teks berupa teks kalimat bahasa Melayu Pontianak yang belum diberikan jeda dan disimpan dalam format *.txt.

3.3.7.2 Pengecekan Aturan Kata Kunci dan Pemenggalan Frasa Sesuai Aturan Kata Kunci

Kalimat yang sudah disimpan dalam format *.txt tadi selanjutnya diperiksa berdasarkan aturan kata kunci dan kemuculan jeda yang telah dibuat dan dipenggal menjadi penggalan frasa jeda ucapan.

3.3.7.3 Output

Output dari proses kata kunci adalah penggalan frasa dari kalimat teks bahasa Melayu Pontianak sesuai aturan kata kunci yang dibuat yang disimpan dalam file format *.txt.

Hasil dari proses ini adalah teks kalimat bahasa Melayu Pontianak beserta prediksi jedanya. Misalnya :

1. “Semue-mue-e/1 tepat waktu .”
2. “Aoklah-aoklah/2 aku ngakulah/1 kalok pemerintah yang tak tepat waktu .”
3. ” Dahlah/2 kite tunggu jak/1 beberape ari agik.”

3.3.8 Implementasi Prediksi Jeda

3.3.8.1 Implementasi Prediksi Jeda Menggunakan Aturan Kata Kunci

Sistem prediksi jeda dimulai dengan memasukkan teks kalimat Bahasa Melayu Pontianak yang kemudian melewati beberapa proses hingga menghasilkan keluaran berupa teks kalimat Bahasa Melayu Pontianak beserta prediksi jedanya. Rancangan antar muka sistem dapat dilihat pada gambar 3.4

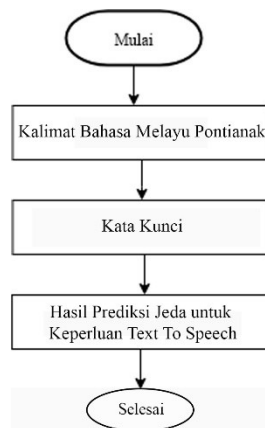
The image shows a web application interface with the following components:

- Label:** A text label at the top left.
- Input Field:** A large text box containing the placeholder text "Input Teks Kalimat Bahasa Melayu Pontianak".
- Button:** A button labeled "Tombol Prediksi" located below the input field.
- Output Field 1:** A text box containing the placeholder text "Hasil Aturan Kata Kunci".
- Output Field 2:** A text box containing the placeholder text "Hasil Prediksi Jeda untuk Keperluan Text To Speech".

Gambar 3.4 Rancangan Antarmuka Sistem Kata Kunci

Antarmuka sistem prediksi jeda berupa halaman web yang berisi sebuah tombol untuk memulai proses prediksi jeda dan tiga kolom untuk menampilkan hasil dari tiap proses prediksi jeda.

Tiap proses mengikuti diagram alir proses sistem prediksi jeda yang dapat dilihat pada gambar 3.5



Gambar 3.5 Diagram Alir Proses Prediksi Jeda Kata Kunci

Proses prediksi jeda dimulai dengan memasukkan teks kalimat Bahasa Melayu Pontianak yang akan diprediksi jeda. Teks masukan tersebut kemudian diolah menggunakan aturan kata kunci untuk mendapatkan jeda. Jeda yang dihasilkan selanjutnya diberi penandaan indeks jeda. Kemudian hasil penandaan indeks jeda di normalisasi menjadi *output* yang diperlukan untuk keperluan sistem TTS.

3.3.8.2.1 Kata Kunci

Hasil Aturan Kata Kunci merupakan proses pemenggalan frasa berdasarkan aturan kata kunci yang telah dibuat beserta ditandai kejadian indeks jeda singkat dan panjang. Hasil dari penandaan jeda disimpan dalam file *.txt yang berisi kalimat bahasa Melayu Pontianak yang telah ditandai jeda singkat (/1) dan jeda panjang (/2).

Contoh kalimat :

1. Semue-mue-e/1 tepat waktu.
2. Aoklah-aoklah/2 aku ngakulah/1 kalok pemerintah yang tak tepat waktu.
3. Dahlah/2 kite tunggu jak/1 beberape ari agik.

3.3.8.2.2 Hasil Keperluan Jeda untuk Sistem *Text To Speech*

Hasil dari penandaan jeda kemudian dinormalisasi dengan mengubah tanda jeda singkat(/1) menjadi tanda “|” dan mengubah tanda jeda panjang(/2) mejadi tanda “||” untuk keperluan implementasi sistem *Text-to-Speech*.

Contoh kalimat:

1. Semue-mue-e| tepat waktu.
2. Aoklah-aoklah|| aku ngakulah| kalok pemerintah yang tak tepat waktu.
3. Dahlah|| kite tunggu jak| beberape ari agik.

3.3.8.2 Implementasi Prediksi Jeda Menggunakan Aturan Kata Kunci PoS

Sistem prediksi jeda dimulai dengan memasukkan teks kalimat Bahasa Melayu Pontianak yang kemudian melewati beberapa proses hingga menghasilkan keluaran berupa teks kalimat Bahasa Melayu Pontianak beserta prediksi jedanya. Rancangan antar muka sistem dapat dilihat pada gambar 3.6

Label

Input Teks Kalimat Bahasa Melayu Pontianak

Tombol Prediksi

Hasil Tagging PoS

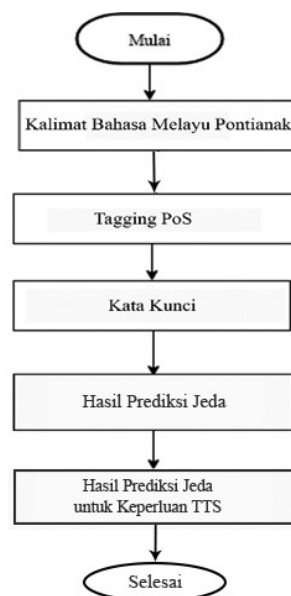
Hasil Penandaan Jeda

Hasil Prediksi Jeda untuk Keperluan Text To Speech

Gambar 3.6 Rancangan Antarmuka Sistem Kata Kunci Menggunakan PoS

Antarmuka sistem prediksi jeda berupa halaman web yang berisi sebuah tombol untuk memulai proses prediksi jeda dan empat kolom untuk menampilkan hasil dari tiap proses prediksi jeda.

Tiap proses mengikuti diagram alir proses sistem prediksi jeda yang dapat dilihat pada gambar 3.7



Gambar 3.7 Diagram Alir Proses Prediksi Jeda Menggunakan PoS

Proses prediksi jeda dimulai dengan memasukkan teks kalimat Bahasa Melayu Pontianak yang akan diprediksi jedanya. Teks masukan tersebut kemudian diolah untuk mendapatkan penambahan spasi dan di *tagging* untuk mendapatkan PoS menggunakan bantuan IPOSTagger. Hasil *tagging* PoS selanjutnya diolah menggunakan kata kunci untuk mendapatkan jeda. Jeda yang dihasilkan selanjutnya diberi penandaan indeks jeda. Kemudian hasil penandaan

indeks jeda di normalisasi menjadi *output* yang diperlukan untuk keperluan sistem TTS

3.3.8.2.1 Tagging PoS

Tagging PoS adalah proses yang dilakukan untuk mendapatkan *tag* PoS atau kelas kata dari setiap kata yang ada pada kalimat masukan. Setelah kalimat mendapatkan penambahan spasi dan titik , kemudian ditandai untuk mendapatkan *tag* PoS dengan menggunakan aplikasi IPOSTagger yang telah dilatih sebelumnya dan hasilnya disimpan pada *file* dengan format *.hsl yang berisikan “kata/PoS”.

Contoh kalimat:

1. Semue-mue-e/NNP tepat/DRB waktu/NNU ./.
2. Aoklah-aoklah/RP aku/PRP ngakulah/VBI kalok/CON pemerintah/NN yang/CON tak/NEG tepat/DRB waktu/NNU ./.
3. Dahlah/UH kite/PRP tunggu/VBI jak/RB beberape/CDI ari/NN agik/RB ./.

3.3.8.2.2 Kata Kunci

Hasil dari *tagging* PoS kemudian diproses menggunakan kata kunci untuk pemenggalan frasa berdasarkan aturan kata kunci yang telah dibuat beserta ditandai kejadian indeks jeda singkat dan panjang. Hasil dari penandaan jeda disimpan dalam file *.txt yang berisi kalimat bahasa Melayu Pontianak yang telah ditandai jeda singkat (/1) dan jeda panjang (/2) beserta label PoS.

Contoh kalimat :

1. Semue-mue-e/NNP/1 tepat/DRB waktu/NNU ./.
2. Aoklah-aoklah/RP/2 aku/PRP ngakulah/VBI/1 kalok/CON pemerintah/NN yang/CON tak/NEG tepat/DRB waktu/NNU ./.
3. Dahlah/UH/2 kite/PRP tunggu/VBI jak/RB/1 beberape/CDI ari/NN agik/RB ./.

3.3.8.2.3 Hasil Prediksi Jeda

Hasil dari kata kunci kemudian dilakukan penghapusan label PoS untuk hasil akhir dari prediksi jeda. Hasil tersebut kemudian disimpan dalam file *.txt yang berisi kalimat bahasa Melayu Pontianak yang telah ditandai jeda singkat (/1)

dan jeda panjang (/2).

1. Semue-mue-e/1 tepat waktu .
2. Aoklah-aoklah/2 aku ngakulah/1 kalok pemerintah yang tak tepat waktu.
3. Dahlah/2 kite tunggu jak/1 beberapa ari agik.

3.3.8.2.4 Hasil Prediksi Jeda untuk Keperluan Sistem *Text To Speech*

Hasil dari penandaan jeda kemudian dinormalisasi dengan mengubah tanda jeda singkat(/1) menjadi tanda “|” dan mengubah tanda jeda panjang(/2) mejadi tanda “||” untuk keperluan implementasi sistem *Text-to-Speech*.

Contoh kalimat:

1. Semue-mue-e| tepat waktu .
2. Aoklah-aoklah|| aku ngakulah| kalok pemerintah yang tak tepat waktu.
3. Dahlah|| kite tunggu jak| beberapa ari agik.

3.3.9 Perancangan Pengujian

Pengujian dilakukan dengan jumlah 500 kalimat dengan aturan kata kunci yang dibandingkan dengan kalimat jeda dari penutur. Pengujian akurasi jeda dilakukan agar didapatkan nilai acuan pada *aturan* kata kunci dalam menghasilkan potongan jeda. Pengujian dilakukan untuk mendapatkan nilai pengujian yang akan dianalisis menjadi kesimpulan. Pengujian dilakukan dengan membandingkan kalimat asli dengan kalimat hasil prediksi sistem menggunakan *Precision*, *Recall* dan *F-Score*. Total kalimat yang akan diuji adalah 500 kalimat korpus. Korpus ini yang akan dibandingkan 500 korpus hasil olahan dari *aturan* kata kunci.

3.3.9.1 Pengujian Kecocokan Frasa Jeda Dalam Satu Kalimat Penuh

Pengujian untuk menguji hasil kecocokan jeda dalam satu kalimat penuh dengan cara membandingkan hasil jeda dari penutur Bahasa Melayu Pontianak dengan hasil jeda pada aturan kata kunci. Jeda terbagi menjadi 2 macam yaitu jeda 2 (panjang) dan jeda 1 (singkat)+2 (panjang) . Akurasi pengujian ini adalah rasio antara jeda yang sesuai dengan jeda penutur dan total jumlah jeda yang benar dengan jeda aturan kata kunci.

Rumus untuk uji akurasi jeda dalam satu kalimat penuh adalah sebagai berikut:

$$\text{Akurasi frasa jeda dalam kalimat} = \frac{\text{jumlah kalimat yang sesuai}}{\text{Total kalimat yang diuji}} \times 100$$

Tabel 3.6 Rencana Pengujian pada Frasa Jeda dalam Satu Kalimat

No .	Jeda Penutur	Jeda Aturan kata kunci	Jeda 1+2		Jeda 2	
			Sesuai	Tidak Sesuai	Sesuai	Tidak Sesuai
			√	X	√	X
Total						

Jumlah kalimat yang diuji adalah sebanyak 500 kalimat dari frasa jeda penutur dan 500 kalimat dari frasa jeda hasil aturan kata kunci. Pengujian dilakukan sebanyak dua kali untuk membandingkan jeda 1+2 dan jeda 2. Total dari kalimat yang sesuai dimasukkan ke dalam persamaan 3.1 untuk mendapatkan hasil akurasi frasa jeda dalam kalimat.

Tabel 3.7 Skenario Pengujian Frasa Ucapan Jeda Panjang pada Satu Kalimat

No.	Jeda Penutur	Jeda Aturan Kata Kunci	Jeda 2	
			Sesuai	Tidak Sesuai
1.	Kalok dah macam tu/2 , mane pulak aku nak menang .	Kalok dah macam tu/2 , mane pulak aku nak menang .	√	X
2.	Ujong-ujong-e/2 , njahat-njahatkan pemerintah .	Ujong-ujong-e/2 , njahat-njahatkan pemerintah .	√	X
N-kalimat	Tempat-e/1 tu luas benar .	Tempat-e tu/1 luas benar .	X	√
Total			2	1

Tabel 3.8 Skenario Pengujian Jeda Singkat+Panjang pada Satu Kalimat

No.	Jeda Penutur	Jeda Aturan Kata Kunci	Jeda 2	
			Sesuai	Tidak Sesuai
1.	Begitu sampai/1 tang luar/2 , ku tengok tukang pos tu/1 barok nak ngidopkan motor.	Begitu sampai/1 tang luar/2 , ku tengok tukang pos tu/1 barok nak ngidopkan motor .	√	X
2.	Mane saye tau bang/2 , saye kan/1 cume njalankan tugas jak .	Mane saye tau bang/2 , saye kan/1 cume njalankan tugas jak .	√	X
N- Kalimat	Ade gak masih sikit aku ni bebar .	Ade gak/1 masih sikit aku ni/1 bebar .	X	√
Total			2	1

Berdasarkan tabel diatas, jeda penutur merupakan jeda pembandingan yang didapat dari penutur, jeda aturan kata kunci merupakan jeda yang akan diuji, *field* sesuai merupakan hasil dari pembandingan antara jeda penutur dan jeda aturan kata kunci yang sama, dan *field* tidak sesuai merupakan hasil pembandingan antara jeda penutur dan jeda aturan kata kunci yang tidak sama. Data uji yang digunakan sebanyak 500 kalimat.

3.3.9.2 Pengujian *Precision*, *Recall* dan *F-Measure*

Pengujian *Recall* dan *Precision* merupakan pengujian untuk mendapatkan informasi hasil pencarian dokumen yang relevan dengan dokumen asli yang ingin dibandingkan. *Precision* dapat dianggap sebagai ukuran ketepatan atau ketelitian, sedangkan *recall* adalah pengulangan kejadian yang sama dengan dokumen asli. Pengujian dilakukan secara berurutan dengan membagi 500 kalimat menjadi 100, 200, 300, 400, dan 500 kalimat. Dari hasil perbandingan tersebut dapat dilakukan perhitungan nilai akurasi *Precision*, *Recall*, dan *F-Measure*. Pegujian

menggunakan 2 kali pengujian yaitu pengujian yang dilakukan menggunakan PoS dan pengujian yang dilakukan tanpa menggunakan PoS. Rancangan pengujian *precision* dan *recall* adalah seperti berikut :

3.3.9.2.1 Skenario Pengujian Jumlah Kalimat

Untuk melihat perbandingan hasil pengujian dibuat skenario jumlah korpus pengujian dengan jumlah korpus 100, 200, 300, 400 dan 500.

Tabel 3.9 Skenario Pengujian Terhadap 100 Kalimat dari 500 Kalimat Korpus Analisis

No .	Kalimat Korpus Uji	Relevan		Tidak Relevan	Precision ($\frac{a}{a+c}$)	Recall ($\frac{a}{a+b}$)	F-Measure ($2 \cdot \frac{p \cdot r}{p+r}$)
		Terambil (a)	Tidak Terambil (b)	Terambil(c)			
1.	0-100						
2.	101-200						
3.	201-300						
4.	301-400						
5.	401-500						

Tabel 3.10

Skenario Pengujian Terhadap 200 Kalimat dari 500 Kalimat Korpus Analisis

No.	Kalimat Korpus Uji	Relevan		Tidak Relevan	Precision ($\frac{a}{a+c}$)	Recall ($\frac{a}{a+b}$)	F-Measure ($2 \cdot \frac{p \cdot r}{p+r}$)
		Terambil (a)	Tidak Terambil (b)	Terambil(c)			
1.	0-200						
2.	0-100, 201-300						
3.	0-100, 301-400						
4.	0-100, 401-500						
5.	101-300						
6.	101-200, 301-400						
7.	101-200, 401-500						
8.	201-400						
9.	201-300, 401-500						
10.	301-500						

Tabel 3.11

Skenario Pengujian Terhadap 300 Kalimat dari 500 Kalimat Korpus Analisis

No .	Kalimat Korpus Uji	Relevan		Tidak Relevan Terambil(c)	Precision ($\frac{a}{a+c}$)	Recall ($\frac{a}{a+b}$)	F- Measure (2. $\frac{p \cdot r}{p+r}$)
		Teram bil (a)	Tidak Terambil (b)				
1.	0-300						
2.	0-200, 301-400						
3.	0-200, 401-500						
4.	0-100, 201-400						
5.	0-100, 201-300, 401-500						
6.	0-100, 301-500						
7.	101-400						
8.	101-300, 401-500						
9.	101-200, 301-500						
10.	201-500						

Tabel 3.12

Skenario Pengujian Terhadap 400 Kalimat dari 500 Kalimat Korpus Analisis

No .	Kalimat Korpus Uji	Relevan		Tidak Relevan Terambil(c)	Precision ($\frac{a}{a+c}$)	Recall ($\frac{a}{a+b}$)	F- Measure (2. $\frac{p \cdot r}{p+r}$)
		Teram bil (a)	Tidak Terambil (b)				
1.	0-400						
2.	0-300, 401-500						
3.	0-200, 301-500						
4.	201-500						
5.	101-400						

Tabel 3.13 Skenario Pengujian Terhadap 500 Kalimat dari 500 Kalimat Korpus Analisis

No .	Kalimat Korpus Uji	Relevan		Tidak Relevan	Precision ($\frac{a}{a+c}$)	Recall ($\frac{a}{a+b}$)	F-Measure ($2 \cdot \frac{p \cdot r}{p+r}$)
		Terambil (a)	Tidak Terambil (b)	Terambil(c)			
1.	0-500						

3.3.9.2.2 Rancangan Pengujian Penggalan Frasa Jeda dalam Satu Kalimat

Pengujian penggalan frasa jeda dalam satu kalimat digunakan untuk melihat persamaan penggalan frasa jeda dalam kalimat asli dan kalimat hasil prediksi aturan kata kunci.

Skenario pengujian kemunculan jeda dapat dilihat pada tabel 3.13.

Tabel 3.14 Skenario Pengujian pada Frasa Jeda dalam Satu Kalimat Untuk Jeda Singkat dan Jeda Panjang

No .	Frasa Penutur	Frasa Aturan Kata Kunci	Relevan		Tidak Relevan	Precision ($\frac{a}{a+c}$)	Recall ($\frac{a}{a+b}$)
			Terambil	Tidak Terambil	Terambil		
1	Contoh Kalimat	Contoh Kalimat	A	b	c	a+c	a+b
Total			A	b	c	a+c	a+b

Keterangan :

a = jumlah frasa jeda / jeda yang relevan ditemukan

b = jumlah frasa jeda / jeda yang relevan tidak ditemukan

c = jumlah frasa jeda / jeda tidak relevan yang ditemukan

Contoh

pengujian terhadap kalimat pada pengujian potongan frasa jeda 1 dapat dilihat pada tabel 3.15 sebagai berikut:

Tabel 3.15 Contoh Skenario Pengujian Potongan Frasa Jeda Panjang dalam Satu Kalimat

No.	Frsa Penutur	Frasa Aturan Kata Kunci	Sesuai		Tidak Sesuai	Precisi on ($\frac{a}{a+c}$)	Recall ($\frac{a}{a+b}$)
			Tera mbil (a)	Tidak Terambil (b)	Terambil(c)		
1.	Aok/1 ingat benar aku .	Aok/1 ingat benar aku .	2	0	0	2	2
2.	Kitak boleh cayak/1 boleh tadak/1 kalok aku sih/1 tak cayak .	Kitak boleh cayak boleh/1 tadak/1 kalok aku sih/1 tak cayak .	2	2	2	4	4
N- kal im at	Tempat-e/1 tu luas benar .	Tempat-e tu/1 luas benar .	0	2	2	2	2
Total			4	4	4	8	8

Pengujian penggalan frasa jeda singkat digunakan untuk melihat persamaan penggalan frasa jeda dari kalimat asli dan kalimat hasil prediksi aturan kata kunci. Pengujian penggalan frasa jeda singkat menggunakan perbandingan penggalan frasa dengan kejadian jeda singkat dan jeda panjang. Total kalimat yang diuji sebanyak 500 kalimat dari korpus analisis dan 500 kalimat dari hasil kata kunci yang akan dipecah sesuai skenario pengujian jumlah kalimat.

Berdasarkan tabel diatas, untuk menghitung *Precision*, yaitu jumlah dokumen sesuai yang terambil adalah a, sedangkan jumlah dokumen yang terambil dalam proses *aturan* kata kunci adalah a+c. Maka rumus *precision* dapat dinyatakan sebagai berikut:

$$Precision = \frac{a}{a+c}$$

Dengan P yang didapatkan dengan membandingkan jumlah frasa jeda yang *sesuai* yang terpanggil dengan jumlah frasa jeda yang terpanggil dalam aturan kata kunci.

Untuk menghitung *recall*, tentukan jumlah dokumen sesuai yang terambil adalah a, sedangkan jumlah dokumen dalam proses aturan kata kunci a+b. Maka rumus *recall* dapat dilihat sebagai berikut:

$$Recall = \frac{a}{a+b}$$

Dengan R adalah *recall*, maka nilai R didapatkan dengan membandingkan farasa jeda *sesuai* yang terpanggil dalam jumlah frasa sesuai yang ada di dapat oleh penutur.

F-Measure merupakan pengukuran yang mengkombinasikan *precision* dan *recall* yang di terapkan kedalam deret harmonic. Rumus *F-Measure* dapat dilihat sebagai berikut:

$$F\text{-measure} = 2 \cdot \frac{precision \cdot recall}{precision + recall}$$

Dengan F adalah *F-Measure*, maka nilai F didapatkan dengan membandingkan *precision* kali *recall* dengan ditambah *recall* kemudian dikali dua.

Pengujian penggalan frasa jeda singkat digunakan untuk melihat persamaan penggalan frasa jeda dari kalimat asli dan kalimat hasil prediksi aturan kata kunci. Pengujian penggalan frasa jeda singkat menggunakan perbandingan penggalan frasa dengan kejadian jeda singkat dan jeda panjang. Total kalimat yang diuji sebanyak 500 kalimat dari korpus analisis dan 500 kalimat dari hasil aturan kata kunci yang akan dipecah sesuai skenario pengujian jumlah korpus pada bagian

3.3.10 Analisis Hasil Pengujian

Pada tahap ini, hasil pengujian sistem akan di analisis secara keseluruhan untuk mempermudah penarikan kesimpulan.

3.3.11 Penarikan Kesimpulan

Kesimpulan dirumuskan berdasarkan tahapan-tahapan yang telah

dilakukan sebelumnya apakah sistem yang dirancang dan dibangun dapat berjalan baik.

BAB IV

Jeda awal	Jeda Akhir	Durasi Jeda	Jumlah Kemunculan
0.6445475	0.6459904	0.001	1
5.3936384	5.3999506	0.006	1
1.4805433	1.4871555	0.007	1
3.2740529	3.2848776	0.011	3
4.4084134	4.4192347	0.011	
4.4432269	4.4543107	0.011	
1.8115121	1.8250113	0.013	
1.986097	1.999496	0.013	3
3.6817244	3.6950429	0.013	1
7.8831156	7.8972757	0.014	
2.7735575	2.7902864	0.017	
3.8444939	3.8614356	0.017	
3.3583633	3.3777758	0.019	2
5.49	5.5094529	0.019	
6.7718082	6.7913141	0.02	
2.9380365	2.95885	0.021	
4.0075	4.03	0.022	1
0.6475	0.67	0.023	1
...

HASIL DAN ANALISIS

4.1 Hasil Penelitian

Hasil dalam penelitian ini adalah berupa pemenggalan – pemenggalan kalimat Bahasa Melayu Pontianak yang berupa frasa kata yang membentuk jeda ucapan yang dibentuk berdasarkan kata kunci.

4.1.1 Hasil Pengkategorian Indeks Jeda dan Penandaan Jeda di Korpus

Nilai "sil" dari *file **'breaks diolah dengan bantuan *spreadsheet* untuk menghitung nilai durasi tiap jeda.

Tabel 4.1 Tabel Perhitungan Durasi Jeda untuk Durasi Indeks Jeda

Nilai durasi jeda digunakan untuk menghitung nilai min , max, median, dan mean dari nilai durasi jeda yang telah dijelaskan pada bagian 3.3.3 . Nilai modus didapat dari nilai yang paling besar di kolom jumlah kemunculan. Hasil perhitungan nilai min, max, median, mean dan modus dapat dilihat pada pada tabel 4.2.

Tabel 4.2 Hasil perhitungan durasi jeda penutur

MEAN	0.333
MEDIAN	0.177
MODUS	0.025
NILAI MIN	0.001
NILAI MAX	2.377

Berdasarkan perhitungan dari kejadian jeda yang terjadi, nilai rata-rata dari seluruh kejadian jeda adalah 0.333 detik. Nilai median atau nilai tengah dari rentang kejadian jeda adalah 0.177 detik. Nilai modus atau nilai yang paling sering muncul adalah 0.025 detik. Nilai terkecil dari kejadian jeda adalah 0.001 detik dan nilai terbesarnya adalah 2.377 detik. Hasil perhitungan dimasukkan ke dalam kolom durasi jeda pada tabel 4.3 untuk kategori indeks jeda.

Tabel 4.3 Kategori Indeks Jeda dengan durasi jeda

Indeks Jeda	Keterangan	Durasi Jeda
0	Tidak ada jeda	0 – < 0.025 s
1	Jeda yang terjadi tipis	0.025 – ≤ 0.33 s
2	Jeda yang terjadi tebal	> 0.33 s
, ! ?	Tanda baca	, ! ?
.	Akhir kalimat	.

Nilai modus digunakan sebagai batas atas untuk lama indeks jeda 0 yang berarti lama kejadian jeda 0 adalah 0 sampai 0.025 detik. Untuk nilai *mean* digunakan sebagai batas atas untuk lama indeks jeda singkat (1) yang memiliki batas 0.025 – 0.333 detik. Untuk jeda panjang (2) menggunakan nilai lebih dari 0.333 detik.

Durasi jeda untuk indeks jeda dari tabel 4.2, kemudian digunakan untuk menandai kejadian jeda penutur di korpus teks bahasa Melayu Pontianak. Hasil penandaan kemunculan jeda sesuai kejadian jeda penutur dapat dilihat pada tabel 4.4

Tabel 4.4 Hasil Penandaan Jeda Pada Korpus

No	Kalimat Bahasa Melayu Pontianak
1.	Kau bikin janji jam limak /2, make/1 jam limak/1 kau haros datang.
2.	Tadak macam tabiat kau tu/2 , kau bikin janji jam limak/2, jam tujuh kau baru datang .
3.	Biarpon betekak macam ape/2, kalok dah soal pemerintah/1 tang negare ini ni/2, dahlah/2, tak usah nak diumungkan ageklah . Tak ade bahase yang tepat nak nerangkan-e .
4.	Pasal-e/2 waktu aku tekanja/1 waktu ade budak kecil masok kamar aku lalu mbangonkan aku/2 yang pertame ku lakukan/1 , ngesat-ek mulot aku pakai guling .
5.	Betol mang kate kawan aku/1 tang Belande tu/1 kalok soal tepat waktu/2, tak ade yang bise dibanggekan dari negara ini.
6.	Belom lama kapal tu ninggalkan Rasau/2 , lalulah/1 dua orang barat tu tepoto-poto .
7.	Duak orang barat tu ngakuk/1 kalok sepanjang jalan dari Rasau ke Kubu/2 , pemandangan alamnye/1 cantek luar biase .
8.	Selamak kamek di Kubu/2 , bukan cume duak orang barat tu jak yang sepok laot/2 kamek semue pon/1 yang orang pribumi sepok gak .
9.	Kamek pon/1 macam yeye ngayalkan utan Amajon/2, ngayalkan pelam Rambo/2 , ngayalkan tengah suting acara petualangan/2, macam tang tepi-tepi .
10.	Aku yang tadak bawak kamera/1 cume bise takjub jak/1 ngan pemandangan yang ade.

Penandaan angka satu dan dua disesuaikan dengan nilai durasi "sil" di *wavesurfer*. Jika nilai durasi "sil" bernilai 0.025 – 0.333 detik, maka "sil" yang terjadi setelah kata yang diucapkan diberi tanda "/1" yang menandakan kejadian jeda singkat. Jika nilai durasi "sil" bernilai batas lebih dari 0.333 detik, maka "sil" yang terjadi setelah kata yang diucapkan diberi tanda "/2" yang menandakan kejadian jeda panjang.

4.1.2 Hasil Pengembangan PoS Bahasa Melayu Pontianak

Hasil dari pengembangan PoS bahasa Melayu Pontianak yaitu tabel PoS dengan jumlah sebanyak 46 set PoS. Set PoS bahasa Melayu Pontianak. Tabel Pengembangan PoS Bahasa Melayu Pontianak bisa dilihat pada tabel 4.5

Tabel 4.5 Set PoS Bahasa Melayu Pontianak

No.	PoS	Deskripsi	Contoh Kata
1	VBT	Verba Transitif (Transitive Verb)	Makai, nenggek, njajah
2	VBI	Verba Intransitif (Intransitive Verb)	Banyak, balek, turun
3	IN	Preposition (Kata Depan)	Di, ke, dari, pade
4	UH	Interjection (Kata Seru)	Oi, woi, alamak
5	RP	Partikel	Pon, lah, jak
6	JJ	Adjective	Kaye, lawar, pandai, budoh
7	OP	Open Parenthesis	({ [
8	CP	Close Parenthesis) }]
9	.	Sentence Terminator	. ! ? ...
10	,	Comma	,
11	SYM	Symbol	# % ^ @
12	MD	Modal	nak, haros
13	NEG	Negation	Bukan, jangan, tadak
14	QT	Quotation	" '
15	WP	WH-Pronoun	Ape, siape, berape
16	DT	Determiner	ini, ni, tu, tuh
17	FW	Foreign Word	Wonderfull, story
18	CDP	Primary Numeral	Satu, duak, tige
19	CDO	Ordinal Numeral	Kesatu, keduak, ketige
20	CDI	Irregular Numeral	Beberape, segale, semue
21	CDC	Collective Numeral	Ratusan, ribuan, pulohan
22	RB	Adverb	Paleng, sementare
23	PRP	Personal Pronoun	Aku, saye, kau, die
24	PRL	Locative Pronoun	Sanak, sinik, situk
25	PRN	Number Pronoun	Satu-satunye, dua-duanye, semue-mue-e
26	NNP	Proper Noun	Eropa, Indonesia, Pontianak
27	NNG	Genitive Common Noun	Bukunye, rumahnye
28	NN	Common Noun	Martabat, Janji

Dari 46 set PoS yang terbentuk ada 28 kelas kata yang sama dengan

penelitian Alfian Farizki (2010) dan M.Iqbal Kamiludin (2017) yaitu VBT, VBI, IN, UH, RP, JJ, OP, CP, tanda baca (.), SYM, MD, NEG, QT, WP, DT, FW, CDP, CDO, CDI, CDC, RB, PRP, PRL, PRN, NNP, NNG dan NN.

Tabel 4.6 Pengembangan Set PoS Baru

No.	PoS	Deskripsi	Contoh Kata
1.	AR	Artikulus	Sang, Si
2.	CON	Konjungsi	Dan, kalok
3.	CR	Currency	Rp, \$
4.	WDT	WH-Determiner	Ape, Siape, Barangsiape
5.	US	Unit Simbol	Gr, Kg, Cm
6.	CDF	Fraction Numeral	Setengah, Seperempat
7.	CDA	Kata Bantu Bilangan	Biji, Ekor, Orang, Buah
8.	WPRB	WH-Adverb	Cemane, Ngape
9.	:	Colon	::
10.	NNC	Countable Common Noun	Buku, Rumah, Karyawan
11.	NNU	Uncountable Common Noun	Aek, Gule, Nasi, Ujan
12.	VBR	Verba Reduplikasi	Jalan-jalan, Potong-potong
13.	VBK	Verba Berkonjungsi	Bersalam-salam, Berputar-putar
14.	SL	Slash	/
15.	DS	Dash	-
16.	TRB	Adverb Of Time	Udah, Belom, Dudok, Sekarang
17.	FRB	Adverb Of Frequency	Jarang, Sering, Kadang-kadang
18.	DRB	Adverb Of Degree	Agak, Hamper, Cukop

Selain itu ada 9 kelas kata yang merujuk pada penelitian Setyaningsih (2017) yaitu AR, CON, :, CR, WDT, US, CDF, CDA, dan WPRB. Kelas kata NNC dan NNU merujuk pada penelitian Adriani (2009). Penambahan kelas kata VBR dan VBK untuk membedakan kelas kata kerja yang mengalami pengulangan dan penambahan imbuhan. Penggantian nama untuk tanda “/” menjadi SL dan tanda “-” menjadi DS untuk mempermudah penandaan kelas kata dan membedakan dengan *slash* atau “/” pemisah kata dan PoS.

4.1.3 Hasil Tagging PoS bahasa Melayu Pontianak

Selanjutnya, teks kalimat bahasa Melayu Pontianak ditandai menggunakan set PoS baru yang sudah dikembangkan. Penandaan atau *tagging*

PoS bertujuan untuk melakukan perbandingan terhadap penggunaan kata kunci tanpa PoS dan terhadap penggunaan kata kunci menggunakan PoS, apakah nilai akurasi akan lebih baik atau lebih buruk. Korpus yang telah diberi label PoS dapat dilihat pada tabel 4.7.

Tabel 4.7 Korpus yang telah diberi label PoS

No.	Tagging POS Bahasa Melayu Pontianak
1.	Kau/PRP bikin/VBT janji/NN jam/NN limak/CDP ,/, make/CON jam/NN limak/CDP kau/PRP haros/MD datang/VBI ./.
2.	NN macam/RP itu/DT ./.
4.	Dahlah/UH kite/PRP tunggu/VBI jak/RP beberapa/CDI ari/NN agik/RB ./.
5.	Aoklah/RP ,/, yang/CON sabar/JJ jak/RP ye/RP ,/, aku/PRP nak/MD op/VBI lok/RB ./.
6.	Biarpon/CON betekak/VBI macam/RP ape/WDT ,/, kalok/CON dah/TRB soal/IN pemerintah/NN tang/IN negare/NN ini/DT ni/DT ,/, dahlah/UH ,/, tak/NEG usah/RB nak/MD diumungkan/VBI ageklah/RB ./.
7.	Pasal-e/NN waktu/NNU aku/PRP tekanjat/VBI waktu/NNU ade/VBT budak/NNC kecil/JJ masok/VBI kamar/NNC aku/PRP lalu/CON mbangonkan/VBT aku/PRP yang/CON pertame/CDO ku/PRP lakukan/VBI ,/, ngesat-ek/VBT mulot/NNC aku/PRP pakai/VBT guling/NNC ./.
8.	Aok/RP ingat/VBI benar/DRB aku/PRP ./.
9.	Bedesut/VBI aku/PRP keluar/VBI kamar/NNC tak/NEG peduli/VBI dengan/CON aek/NNU lior/NNU aku/PRP yang/CON masih/TRB melekat/VBI tang/IN mulot/NNC ./.
10.	Begitu/CON sampai/VBT tang/IN luar/PRL ./.

Tiap kata dalam kalimat bahasa Melayu Pontianak diberi tanda “/” atau

garis miring dan diikuti label PoS dari tabel PoS di tabel 4.5 . Penandaan dilakukan secara manual berdasarkan makna dari kata tersebut ketika diucapkan dan dari Kamus Besar Bahasa Indonesia (KBBI). Hasil dari penandaan PoS ini yang digunakan sebagai korpus latih yang disimpan di dalam *subfolder* “resource” pada *folder* IPOSTagger dengan ekstensi .crp. Kemudian proses pelatihan dilakukan melalui *command prompt* pada *folder* IPOSTagger seperti yang terlihat pada gambar 4.1.

Gambar 4.1 Proses Pelatihan IPOSTagger

Proses pelatihan korpus pada IPOSTagger menggunakan perintah “java ipostagger train” pada *command prompt* untuk melatih IPOSTagger sehingga bisa digunakan untuk men-*tagging* kalimat bahasa Melayu Pontianak.

4.1.4 Hasil Aturan Kata Kunci

Kalimat Bahasa Melayu Pontianak yang sudah ditandai jeda dianalisis kejadian jeda sesuai jeda penutur pada korpus jeda yang terjadi pada setiap penggalan frasa sehingga dihasilkan aturan kata kunci. Aturan kata kunci yang dihasilkan digunakan untuk proses pemenggalan frasa untuk mendapatkan frasa jeda. Berdasarkan analisis yang telah dilakukan terdapat 122 aturan kata kunci dari 217 analisis kata kunci tanpa PoS dan 126 aturan kata kunci dari 223 analisis kata kunci menggunakan PoS melalui 500 kalimat. Tabel aturan kata kunci Bahasa Melayu Pontianak seperti pada table seperti berikut:

Tabel 4.8 Analisis Kemunculan Jeda pada Aturan Kata Kunci

No.	Kata Kunci	Jumlah Kemunculan Kata Setelah Jeda	Jumlah Kemunculan Kata Sebelum Jeda	Jumlah Kemunculan Kata Tidak Ada Jeda	Jeda 1	Jeda 2
1.	kalok	22		18	17	5
2.	gak		19	16	19	
3.	gak-e		1		1	

4.	aher-e		1			1
5.	tak	17		39	14	3
6.	dah	3	2	17	5	
7.	dahlah		1			1
8.	itu	1	8	6	6	3
9.	tu	11	48	25	41	8
10.	tang	63		26	61	2

Berdasarkan tabel 4.8 analisis kemunculan jeda aturan kata kunci diatas dijelaskan bahwa kata “gak” terjadi pada setelah jeda karena berdasarkan analisis jeda setelah kata “gak” terjadi 19 kata dan tidak ada jeda terjadi 16 kata maka kata gak terjadi pada setelah jeda, sedangkan untuk jeda 1 dan jeda 2 dilihat dari jeda yang sering muncul dan dipilih jeda terbanyak. Tabel analisa aturan kata kunci tanpa PoS bisa dilihat pada lampiran A.

Tabel 4.9 Aturan Kata Kunci

No.	Kata Kunci	Kemunculan Kata Sesudah Jeda	Kemunculan Kata Sesudah Jeda
1.	kalok	/1	
2.	gak		/1
3.	gak-e		/1
4.	aher-e		/2
5.	dahlah		/2
6.	itu		/1
7.	tuh		/1
8.	tu		/1
8.	tang	/1	
9.	ni		/1
10.	jak		/1

Berdasarkan tabel 4.9 analisis kemunculan jeda kata kunci, proses analisis yang telah dilakukan kata ”kalok” terjadi sebelum jeda dan mendapatkan kemunculan jeda singkat (1), dan kata “gak” terjadi sesudah jeda dan mendapatkan kemunculan jeda singkat (1). Tabel aturan kata kunci tanpa PoS bisa dilihat lebih lengkap pada Lampiran C.

Tabel 4.10 Analisis Kemunculan Jeda Pada Aturan Kata Kunci Menggunakan PoS

No.	Kata Kunci	Jumlah Kemunculan Kata Setelah Jeda	Jumlah Kemunculan Kata Sebelum Jeda	Jumlah Kemunculan Kata Tidak Ada Jeda	POS
1.	tak	17		39	NEG
2.	bukan	7		6	
3.	tadak	8		7	
4.	ndak		6	3	
5.	dah	3	2	17	TRB
6.	sekarang		2	2	
7.	dulok	1	2	1	
8.	dolok		1		
9.	udah	2	2	13	
10.	selamak		1	2	UH
11.	dahlah		1		
12.	woi		1		
13.	eh		1		
14.	yarabbih		1		
15.	e-eh		1		

Berdasarkan tabel 4.10 analisis kemunculan jeda pada aturan kata kunci menggunakan PoS diatas dijelaskan bahwa kata “tadak” terjadi pada setelah jeda karena berdasarkan analisis, kata “tadak” terjadi setelah jeda 8 kata dan tidak ada jeda terjadi 7 kata maka kata tadak terjadi pada setelah jeda dan berdasarkan kelas kata nya kata “tadak” masuk dalam PoS NEG. Set PoS bisa dilihat pada bagian 4.5. Tabel analisa aturan kata kunci menggunakan PoS lebih lengkap bisa dilihat pada lampiran B.

Tabel 4.11 Aturan Kata Kunci Menggunakan PoS

No.	Kata Kunci	Kemunculan Kata Setelah Jeda	Kemunculan Kata Sebelum Jeda	PoS
1.	bukan	/1		NEG

2.	tadak	/1		TRB
3.	ndak		/1	
4.	dulok		/1	
5.	dolok		/1	
6.	dahlah		/2	UH
7.	woi		/1	
8.	Eh		/1	
9.	yarabbih		/1	
10.	e-eh		/2	

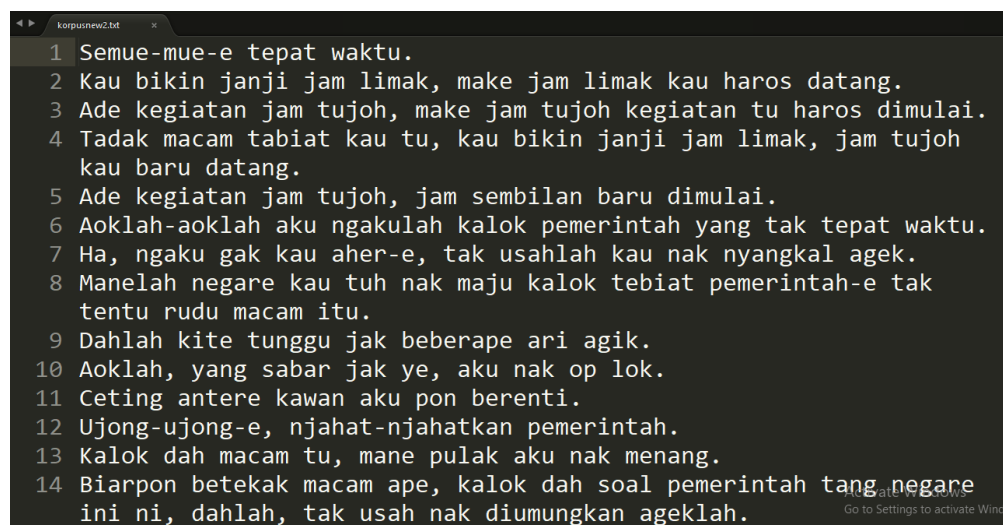
Berdasarkan tabel 4.11 analisis kemunculan jeda aturan kata kunci menggunakan PoS, proses analisis yang telah dilakukan kata “bukan” terjadi sebelum jeda, mendapatkan kemunculan jeda singkat (1) , dan tipe PoS NEG, sedangkan kata “dahlah” terjadi setelah jeda, mendapatkan kemunculan jeda 2 (panjang) dan tipe PoS UH. Set PoS bisa dilihat pada bagian 4.5. Tabel aturan kata kunci tanpa PoS bisa dilihat lebih lengkap pada Lampiran D.

4.1.5 Implementasi Prediksi Jeda

4.1.6.1 Implementasi Aturan Kata Kunci

Hasil penelitian ini yaitu memotong kalimat menjadi frasa berdasarkan aturan kata kunci yang sudah di jelaskan pada 4.8.

Proses aturan kata kunci dilakukan seperti yang telah dijelaskan pada bagian 3.3.7. Inputan untuk proses aturan kata kunci berupa file korpus bahasa Melayu Pontianak yang Pontianak yang belum diberikan jeda dan dismpn dalam format *.txt.



```

1 Semue-mue-e tepat waktu.
2 Kau bikin janji jam limak, make jam limak kau haros datang.
3 Ade kegiatan jam tujuh, make jam tujuh kegiatan tu haros dimulai.
4 Tadak macam tabiat kau tu, kau bikin janji jam limak, jam tujuh
  kau baru datang.
5 Ade kegiatan jam tujuh, jam sembilan baru dimulai.
6 Aoklah-aoklah aku ngakulah kalok pemerintah yang tak tepat waktu.
7 Ha, ngaku gak kau aher-e, tak usahlah kau nak nyangkal agek.
8 Manelah negare kau tuh nak maju kalok tebiat pemerintah-e tak
  tentu rudu macam itu.
9 Dahlah kite tunggu jak beberape ari agik.
10 Aoklah, yang sabar jak ye, aku nak op lok.
11 Ceting antere kawan aku pon berenti.
12 Ujong-ujong-e, njahat-njahatkan pemerintah.
13 Kalok dah macam tu, mane pulak aku nak menang.
14 Biarpon betekak macam ape, kalok dah soal pemerintah tang negare
  ini ni, dahlah, tak usah nak diumungkan ageklah.

```

Gambar 4.2 Teks kalimat bahasa Melayu Pontianak untuk Kata Kunci

Inputan untuk proses kata kunci berupa file korpus bahasa Melayu Pontianak yang disimpan dalam format *.txt.

4.1.6.2.1 Pengecekan dan Pemenggalan Frasa Sesuai Aturan Kata Kunci

Tiap kata dalam kalimat inputan kemudian diperiksa menggunakan aturan kata kunci. Aturan kata kunci yang telah dibuat dapat dilihat pada kode program 4.1.

Kode program 4.1 Aturan kata kunci

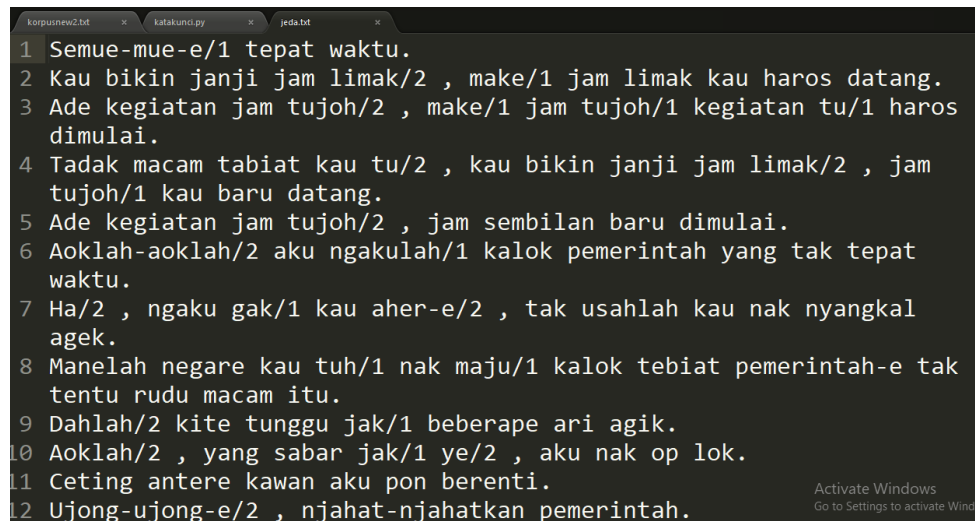
1. with open("korpusnew2.txt") as f1:
2. jsb1 = ['semue-mue-e', 'make', 'tujoh', 'aok', 'gak', 'gak-e', 'itu', 'tuh', 'tu',
'ni', 'jak', 'tapi', 'pasal-e', 'lalu', 'lalulah', 'bebueh-bueh', 'nyemprot-
ek', 'sampai-sampai', 'kan', 'lah', 'woi', 'rupe-e', 'cubelah', 'dulok',
'jalan-jalan', 'sanak', 'agek', 'idop', 'masarakat-e', 'supaye', 'ba', 'agik',
'bahkan', 'puncak-e', 'sepeda-e', 'sikit-sikit', 'ngesak-e', 'rase-e',
'dolok', 'jadilah', 'eh', 'pokok-e', 'njawab-e', 'sebenar-e', 'kayak-e',
'bilang', 'nebang-ek', 'mbangon-ek', 'sengeh', 'peratoran', 'ndak',
'muleilah', 'jangankan', 'yarabbih', 'yak', 'nurotlah', 'negare-e',
'masalah', 'sergio', 'jage', '%', 'awal-e', 'mbayangkan', 'atau',
'angkot', 'kujelaskan', 'sejarah-e', 'philibert', 'tulong', 'kabarkan',
'tanggal', 'sikit-sikit', 'momen', 'salah', 'pagi', 'karne'] + tokens
3. jst1 = ['kalok', 'tang', 'kamek', 'sambel', 'ketemu', 'bukan', 'pas', 'tadak',
'cume', 'palak', 'daripade', 'pasti', 'timbol', 'antare', 'atau', 'teros',
'terdengar', 'nuju']
4. jsb2 = ['aoklah-aoklah', 'aher-e', 'dahlah', 'sih', 'bang', 'tibe-tibe', 'name-e',
'kate-e', 'bah', 'rakyat-e', 'sd', 'tegal-e', 'membelah-ek', 'kerje-e',
'rupiah', 'e-eh', 'orange-nassau', 'spanyol', 'beatrix', 'kajol',
'ngambek-ek', 'lumayanlah']
5. jst2 = ['kire-kire', 'kupoto-ek']
6. fileresult = open("jeda.txt", "w+")

Kalimat akan dipenggal berdasarkan kata kunci yang telah dibuat, untuk jst1 yaitu kata setelah jeda dengan kemunculan jeda 1 (singkat), jsb1 yaitu kata sebelum jeda dengan kemunculan jeda 1 (singkat), jst2 yaitu kata setelah jeda

dengan kemunculan jeda 2 (panjang), jsb2 yaitu kata sebelum jeda dengan kemunculan jeda 2.

4.1.6.2.2 Output

Hasil dari pemenggalan aturan kata kunci disimpan dalam file format *.txt dan merupakan penggalan frasa dari kalimat teks bahasa Melayu Pontianak sesuai aturan kata kunci yang dibuat. Output dapat dilihat pada gambar 4.3



```

korpusnew2.txt x katakunci.py x jeda.txt x
1 Semue-mue-e/1 tepat waktu.
2 Kau bikin janji jam limak/2 , make/1 jam limak kau haros datang.
3 Ade kegiatan jam tujuh/2 , make/1 jam tujuh/1 kegiatan tu/1 haros
  dimulai.
4 Tadak macam tabiat kau tu/2 , kau bikin janji jam limak/2 , jam
  tujuh/1 kau baru datang.
5 Ade kegiatan jam tujuh/2 , jam sembilan baru dimulai.
6 Aoklah-aoklah/2 aku ngakulah/1 kalok pemerintah yang tak tepat
  waktu.
7 Ha/2 , ngaku gak/1 kau aher-e/2 , tak usahlah kau nak nyangkal
  agek.
8 Manelah negare kau tuh/1 nak maju/1 kalok tebiat pemerintah-e tak
  tentu rudu macam itu.
9 Dahlah/2 kite tunggu jak/1 beberape ari agik.
10 Aoklah/2 , yang sabar jak/1 ye/2 , aku nak op lok.
11 Ceting antere kawan aku pon berenti.
12 Ujong-ujong-e/2 , njahat-njahatkan pemerintah.
  
```

Gambar 4.3 Output Hasil Aturan Kata Kunci

Output dari aturan kata kunci adalah hasil dari penandaan jeda beserta kemunculan jeda yang terjadi.

Tabel 4.12 Perbandingan jeda penutur dan jeda kata kunci

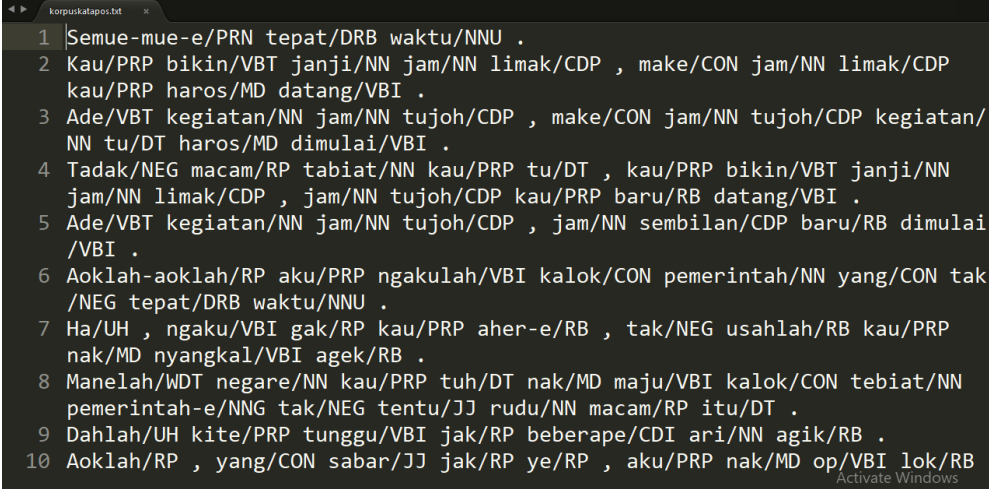
No.	Jeda Penutur	Jeda Aturan Kata Kunci
1.	Begitu sampai/1 tang luar/2 , ku tengok tukang pos tu/1 barok nak ngidopkan motor .	Begitu sampai/1 tang luar/2 , ku tengok tukang pos tu/1 barok nak ngidopkan motor .
2.	Selamak kamek di Kubu/2 , bukan cume duak orang barat tu jak yang sepok laot/2 kamek semue pon/1 yang orang pribumi sepok gak .	Selamak/1 kamek di Kubu/2 , bukan/1 cume duak orang barat tu/1 jak/1 yang sepok laot/1 kamek semue pon yang orang pribumi sepok gak .
3.	Kamek pon/1 macam yeye ngayalkan utan Amajon/2 , ngayalkan pelam Rambo/2 , ngayalkan/1 tengah suting acara petualangan/2 , macam tang tepi-tepi .	Kamek pon macam yeye ngayalkan utan Amajon/2 , ngayalkan pelam Rambo/2 , ngayalkan tengah suting acara petualangan/2 , macam/1 tang tepi-tepi .

Berdasarkan tabel 4.12, penggalan frasa pada kalimat nomor 1 sesuai. Potongan frasa kalimat nomor 2 dan 3 tidak sesuai, karena pada jeda penutur, kata selamak tidak memiliki jeda sedangkan jeda lookup tabel memiliki jeda dikarenakan berdasarkan analisis kata selamak lebih banyak setelah jeda daripada tidak ada jeda.

4.1.5.2 Implementasi Aturan Kata Kunci Menggunakan PoS

Hasil penelitian ini yaitu memotong kalimat menjadi frasa berdasarkan aturan kata kunci yang sudah di jelaskan pada 4.14.

Proses aturan kata kunci dilakukan seperti yang telah dijelaskan pada bagian 3.3.7 di mulai dari membaca inputan kalimat Bahasa Melayu Pontianak.



```

1 Semue-mue-e/PRN tepat/DRB waktu/NNU .
2 Kau/PRP bikin/VBT janji/NN jam/NN limak/CDP , make/CON jam/NN limak/CDP
  kau/PRP haros/MD datang/VBI .
3 Ade/VBT kegiatan/NN jam/NN tujuh/CDP , make/CON jam/NN tujuh/CDP kegiatan/
  NN tu/DT haros/MD dimulai/VBI .
4 Tadak/NEG macam/RP tabiat/NN kau/PRP tu/DT , kau/PRP bikin/VBT janji/NN
  jam/NN limak/CDP , jam/NN tujuh/CDP kau/PRP baru/RB datang/VBI .
5 Ade/VBT kegiatan/NN jam/NN tujuh/CDP , jam/NN sembilan/CDP baru/RB dimulai
  /VBI .
6 Aoklah-aoklah/RP aku/PRP ngakulah/VBI kalok/CON pemerintah/NN yang/CON tak
  /NEG tepat/DRB waktu/NNU .
7 Ha/UH , ngaku/VBI gak/RP kau/PRP aher-e/RB , tak/NEG usahlah/RB kau/PRP
  nak/MD nyangkal/VBI agek/RB .
8 Manelah/WDT negare/NN kau/PRP tuh/DT nak/MD maju/VBI kalok/CON tebiat/NN
  pemerintah-e/NNG tak/NEG tentu/JJ rudu/NN macam/RP itu/DT .
9 Dahlah/UH kite/PRP tunggu/VBI jak/RP beberape/CDI ari/NN agik/RB .
10 Aoklah/RP , yang/CON sabar/JJ jak/RP ye/RP , aku/PRP nak/MD op/VBI lok/RB
  
```

Gambar 4.4 Text Kalimat Bahasa Melayu Pontianak untuk Aturan Kata Kunci Menggunakan PoS

Inputan untuk proses kata kunci berupa file korpus bahasa Melayu Pontianak yang sudah diberi label PoS dan disimpan dalam format *.txt.

4.1.6.2.3 Pengecekan dan Pemenggelan Frasa Jeda Menggunakan PoS

Tiap kata dalam kalimat inputan kemudian diperiksa menggunakan aturan kata kunci menggunakan PoS . Aturan kata kunci yang telah dibuat dapat dilihat pada kode program 4.2.

Kode program 4.2 Aturan Kata Kunci Menggunakan PoS

1. `with open("korpusnew2.txt") as f1:`
2. `jsb1 = ['semue-mue-e/prn','tujuh/cdp', 'make/con', 'tapi/con', 'lalu/con',`
`'lalulah/con', 'supaye/con', 'bahkan/con', 'jadi/con', 'tegal-e/con',`
`'karne/con', 'atau/con', 'aok/rp', 'gak/rp', 'gak-e/rp', 'jak/rp',`
`'kan/rp', 'lah/rp', 'ba/rp', 'kayak-e/rp', 'yak/rp', 'lok/rp', 'rupe-e/rb',`
`'agek/rb', 'agik/rb', 'pokok-e/rb', 'sebenar-e/rb', 'biase/rb',`
`'misal-e/rb',`
`'selanjot-e/rb', 'ndak/neg', 'dulok/trb', 'dolok/trb', 'woi/uh', 'eh/uh',`
`'yarabbih/uh', 'itu/dt', 'tuh/dt', 'tu/dt', 'ni/dt', 'jangankan/in',`
`'pasal-e/nn','peratoran/nn', 'masalah/nn', 'tanggal/nn', 'momen/nn',`

'pagi/nn', 'nyemprot-ek/vbt', 'jadilah/vbt', 'nebang-ek/vbt',
 'mbangonek/vbt', 'muleilah/vbt', 'jage/vbt', 'awal-e/vbt',
 'mbayangkan/vbt', 'tulong/vbt', 'kabarkan/vbt', 'diperingatek/vbt',
 'ngambek-ek/vbt', 'ngadap-ek', 'bebueh-bueh/vbk', 'jalan-jalan/vbr',
 'siap-siaplah/vbr', 'sergio/nnp', 'philibert/nnp', 'desember/nnp',
 'cubelah/vbi', 'idop/vbi', 'njawab-e/vbi', 'bilang/vbi', 'angkot/vbi',
 'kujelaskan/vbi', 'sanak/prl', 'tengah-tengah/prl', 'masarakat-e/nng',
 'puncak-e/nng', 'rase-e/nng', 'ngeare-e/nng', 'sejarah-e/nng', 'sikit-
 sikit/jj', 'ngesak-e/jj', 'sengeh/jj', 'nurotlah/jj', 'salah/jj', '%/sym'] +
 tokens

3. **jst1** = ['kalok/con', 'antare/con', 'atau/con', 'sambel/rp', 'daripade/rp',
 'aher- e/rb', 'pasti/rb', 'bukan/neg', 'tadak/neg', 'tang/in', 'kamek/prp',
 'kupoto-ek/vbt', 'nuju/vbt', 'kamera/nnc', 'palak/nnc', 'ketemu/vbi',
 'timbol/vbi', 'terdengar/vbi', 'pas/jj', 'teros/frb']
4. **jsb2** = ['aoklah-aoklah/rp', 'sih/rp', 'bah/rp', 'tibe-tibe/rb', 'bile/rb',
 'dahlah/uh',
 'e-eh/uh', 'bang/nnc', 'sd/nnp', 'orange-nassau/nnp', 'spanyol/nnp',
 'beatrix/nnp', 'kajol/nnp', 'name-e/nng', 'kate-e/nng', 'rakyat-e/nng',
 'sepeda-e/nng', 'kerje-e/nng', 'lumayanlah/jj', 'rupiah/cda']
5. **jst2** = ['membelah-ek/vbt', 'kire-kire/drbr', 'waktu/nnu']
6. **fileresult** = `open("jedapos.txt", "w+")`

Kalimat akan dipenggal berdasarkan kata kunci yang telah dibuat, untuk jst1 yaitu kata setelah jeda dengan kemunculan jeda 1 (singkat), jsb1 yaitu kata sebelum jeda dengan kemunculan jeda 1 (singkat), jst2 yaitu kata setelah jeda dengan kemunculan jeda 2 (panjang), jsb2 yaitu kata sebelum jeda dengan kemunculan jeda 2.

```

1 Semue-mue-e/PRN/1 tepat/DRB waktu/NNU .
2 Kau/PRP bikin/VBT janji/NN jam/NN limak/CDP/2 , make/CON/1 jam/NN limak/
  CDP kau/PRP haros/MD datang/VBI .
3 Ade/VBT kegiatan/NN jam/NN tujuh/CDP/2 , make/CON/1 jam/NN tujuh/CDP/1
  kegiatan/NN tu/DT/1 haros/MD dimulai/VBI .
4 Tadak/NEG macam/RP tabiat/NN kau/PRP tu/DT/2 , kau/PRP bikin/VBT janji/NN
  jam/NN limak/CDP/2 , jam/NN tujuh/CDP/1 kau/PRP baru/RB datang/VBI .
5 Ade/VBT kegiatan/NN jam/NN tujuh/CDP/2 , jam/NN sembilan/CDP baru/RB
  dimulai/VBI .
6 Aoklah-aoklah/RP/2 aku/PRP ngakulah/VBI kalok/CON pemerintah/NN yang/CON
  tak/NEG tepat/DRB waktu/NNU .
7 Ha/UH/2 , ngaku/VBI gak/RP/1 kau/PRP aher-e/RB/2 , tak/NEG usahlah/RB kau/
  PRP nak/MD nyangkal/VBI agek/RB/1 .
8 Manelah/WDT negare/NN kau/PRP tuh/DT/1 nak/MD maju/VBI kalok/CON tebiat/
  NN pemerintah-e/NNG tak/NEG tentu/JJ rudu/NN macam/RP itu/DT/1 .
9 Dahlah/UH/2 kite/PRP tunggu/VBI jak/RP/1 beberape/CDI ari/NN agik/RB/1 .
10 Aoklah/RP/2 , yang/CON sabar/JJ jak/RP/1 ye/RP/2 , aku/PRP nak/MD op/VBI
    lok/RB .

```

Gambar 4.5 Hasil Penandaan Jeda Menggunakan PoS

Hasil dari penghapusan PoS pada Kalimat pada gambar 5.4 selanjutnya diproses pada tahap selanjutnya.

4.1.6.2.4 Penghapusan Label PoS pada Kalimat

Pada bagian ini, kalimat yang sudah berlabel PoS akan dihilangkan. Penghapusan label PoS telah dibuat pada kode program 4.3

Kode program 4.3 Penghapusan Label PoS

1. `with open ('jedapos.txt', 'r') as file :`
2. `filedata= file.read()`
3. `filedata = filedata.replace('/VBR', '')`
4. `filedata = filedata.replace('/VBK', '')`
5. `filedata = filedata.replace('/VBT', '')`
6. `filedata = filedata.replace('/VBI', '')`
7. `with open('prediksi.txt', 'w') as file:`

4.1.6.2.5 Output

Hasil dari pemenggalan aturan kata kunci disimpan dalam file format *.txt dan merupakan penggalan frasa dari kalimat teks bahasa Melayu Pontianak sesuai aturan kata kunci yang dibuat. Output dapat dilihat pada gambar 4.3

```

1 Semue-mue-e/1 tepat waktu .
2 Kau bikin janji jam limak/2 , make/1 jam limak kau haros datang .
3 Ade kegiatan jam tujuh/2 , make/1 jam tujuh/1 kegiatan tu/1 haros dimulai .
4 Tadak macam tabiat kau tu/2 , kau bikin janji jam limak/2 , jam tujuh/1 kau baru
  datang .
5 Ade kegiatan jam tujuh/2 , jam sembilan baru dimulai .
6 Aoklah-aoklah/2 aku ngakulah kalok pemerintah yang tak tepat waktu .
7 Ha/2 , ngaku gak/1 kau aher-e/2 , tak usahlah kau nak nyangkal agek/1 .
8 Manelah negare kau tuh/1 nak maju kalok tebiat pemerintah-e tak tentu rudu macam
  itu/1 .
9 Dahlah/2 kite tunggu jak/1 beberape ari agik/1 .
10 Aoklah/2 , yang sabar jak/1 ye/2 , aku nak op lok .
11 Ceting antere kawan aku pon berenti .
12 Ujong-ujong-e/2 , njahat-njahatkan pemerintah .
13 Kalok dah macam tu/2 , mane pulak aku nak menang .
14 Biarpon betekak macam ape/2 , kalok dah soal pemerintah tang negare ini ni/2 ,
  dahlah/2 , tak usah nak diumungkan ageklah .
15 Tak ade bahase yang tepat nak nerangkan-e .
16 Ade gak/1 yang nyebotnye peker .
17 Same tak iye jak/1 pekong ngan peker tu/1 .

```

Gambar 4.6 Hasil Output Prediksi Jeda Aturan Kata Kunci Menggunakan PoS

Output dari aturan kata kunci adalah hasil dari penandaan jeda beserta kemunculan jeda yang terjadi.

Tabel 4.13 Perbandingan jeda penutur dan jeda kata kunci menggunakan PoS

No.	Jeda Penutur	Jeda Aturan Kata Kunci
1.	Semue-mue-e/1 tepat waktu .	Semue-mue-e/1 tepat waktu .
2.	Selamak kamek di Kubu/2 , bukan cume duak orang barat tu jak yang sepok laot/2 kamek semue pon/1 yang orang pribumi sepok gak .	Selamak kamek di Kubu/2 , bukan cume duak orang barat tu/1 jak/1 yang sepok laot kamek semue pon yang orang pribumi sepok gak/1 ..
3.	Kamek pon/1 macam yeye ngayalkan utan Amajon/2 , ngayalkan pelam Rambo/2 , ngayalkan/1 tengah suting acara petualangan/2 , macam tang tepi- tepi .	Kamek pon macam yeye ngayalkan utan Amajon/2 , ngayalkan pelam Rambo/2 , ngayalkan tengah suting acara petualangan/2 , macam tang tepi-tepi .

Berdasarkan tabel 4.13, penggalan frasa pada kalimat nomor 1 sesuai. Potongan frasa kalimat nomor 2 dan 3 tidak sesuai, karena pada jeda penutur, kata selamak tidak memiliki jeda sedangkan jeda aturan kata kunci memiliki jeda dikarenakan berdasarkan analisis kata selamak lebih banyak setelah jeda daripada tidak ada jeda.

4.1.6 Hasil Implementasi Prediksi Jeda

4.1.6.1 Hasil Implementasi Prediksi Jeda Menggunakan Aturan Kata Kunci

Prediksi jeda ditampilkan dalam halaman *web* untuk memudahkan pengguna melakukan prediksi jeda. Antarmuka dari prediksi jeda dapat dilihat pada gambar 4.7

Gambar 4.7
Antarmuka Sistem
Prediksi Jeda Kata
Pada
Gambar 4.7, dapat
terdapat tiga
kolom dan
tombol pada
web prediksi
Kolom
berfungsi

Antarmuka
Prediksi
Kunci
gambar
dilihat
buah
sebuah
halaman
jeda.
pertama
sebagai

tempat untuk memasukkan teks kalimat bahasa Melayu Pontianak yang ingin diprediksi jeda. Tombol prediksi berfungsi untuk memulai proses prediksi jeda. Kolom kedua berfungsi untuk menampilkan hasil dari proses penandaan jeda berdasarkan aturan kata kunci. Kolom terakhir berfungsi untuk menampilkan hasil akhir dari prediksi jeda berupa teks kalimat yang dapat digunakan untuk keperluan implementasi *text-to-speech*.

Proses prediksi jeda dimulai dari memasukkan teks kalimat bahasa Melayu Pontianak kemudian menekan tombol Prediksi. Selanjutnya tiap proses

menghasilkan *output* seperti yang dijelaskan pada poin berikut ini.

4.1.6.2.1 Kata Kunci

Kalimat dimasukan kemudian diproses menggunakan aturan kata kunci. Pada tahap ini, aturan kata kunci diperiksa apakah ada kata yang memiliki jeda atau tidak sehingga hasil akhir proses berupa teks kalimat bahasa Melayu Pontianak dengan penandaan jeda singkat dan jeda panjang seperti pada gambar

dibawah ini.

Gambar 4.8 Tampilan *output* proses aturan kata kunci

Hasil aturan kata kunci di kolom kedua di proses untuk ditandai jeda singkat dan jeda panjang sehingga hasil akhir berupa teks kalimat bahasa Melayu Pontianak dengan penandaan jeda singkat dan jeda panjang.

4.1.6.2.2 Hasil Prediksi Jeda untuk Keperluan TTS

Pada tahap ini dilakukan normalisasi hasil prediksi jeda untuk mengubah tanda jeda singkat(/1) menjadi tanda “|” dan mengubah tanda jeda panjang (/2) menjadi tanda “||” untuk keperluan implementasi sistem *Text-to-Speech*.

Kode program 4.4 Normalisasi Hasil Jeda

1. `with open('kal_prediksi_jeda.txt', 'r') as file :`
2. `filedata = file.read()`


```

3.         filedata = filedata.replace('/1', '|')
4.         filedata = filedata.replace('/2',' ||')
5.         filedata = filedata.replace('/2',' ||')

6.     with open('kal_prediksi_tts.txt', 'w+') as file:
7.         file.write(filedata)

```

Kode program 4.4 dipakai untuk normalisasi hasil jeda sehingga hasil akhir dari prediksi jeda berupa teks kalimat bahasa Melayu Pontianak dengan tanda baca | dan || untuk jeda. Hasil akhir dapat dilihat pada gambar dibawah ini.

Gambar 4.9 Tampilan *output* hasil akhir prediksi jeda

Hasil prediksi jeda pada kolom keempat kemudian dinormalisasi menjadi hasil akhir prediksi jeda yang digunakan untuk kerpeluan implementasi sistem TTS seperti pada kolom ketiga.

4.1.6.2 Hasil Implemntasi Jeda Aturan Kata Kunci Menggunakan PoS

Prediksi jeda ditampilkan dalam halaman *web* untuk memudahkan pengguna melakukan prediksi jeda. Antarmuka dari prediksi jeda dapat dilihat pada gambar 4.10

Gambar 4.10

Sistem
Kata Kunci
Pada
dapat dilihat
empat buah
sebuah
halaman web
Kolom
berfungsi
tempat untuk
teks kalimat
Melayu
yang ingin
jedanya.
prediksi

Antarmuka
Prediksi Jeda
PoS
gambar 4.7,
terdapat
kolom dan
tombol pada
prediksi jeda.
pertama
sebagai
memasukkan
bahasa
Pontianak
diprediksi
Tombol
berfungsi

untuk memulai proses prediksi jeda. Kolom kedua berfungsi untuk menampilkan hasil dari proses *tagging* PoS. Kolom ketiga untuk menampilkan hasil dari proses penandaan jeda berdasarkan aturan kata kunci. Kolom keempat berfungsi untuk menampilkan hasil akhir dari prediksi jeda dan kolom kelima hasil prediksi jeda berupa teks kalimat yang dapat digunakan untuk keperluan implementasi *text-to-speech*.

Proses prediksi jeda dimulai dari memasukkan teks kalimat bahasa Melayu Pontianak kemudian menekan tombol Prediksi. Selanjutnya tiap proses menghasilkan *output* seperti yang dijelaskan pada poin berikut ini.

4.1.6.2.1 Tagging PoS

Kalimat masukan diolah untuk dilakukan penambahan spasi pada penulisan tanda koma (,), tanda titik(.), tanda seru (!) dan tanda tanya(?). Penambahan spasi dilakukan untuk kalimat yang diketik tanpa menambahkan spasi pada tanda baca, misalnya : “Pergi lah sana.” diubah menjadi “Pergi lah sana .”. Hal ini diperlukan untuk keperluan *tagging* PoS pada IPOSTagger yang

mengenali tanda baca sebagai label PoS juga. Penambahan spasi pada pada kalimat dapat dilihat pada kode program 4.5.

Kode program 4.5 Penambahan spasi pada kalimat

```

1.  def simpan_file(kalimat):
2.      tokens = [',', '.', '!', '?']
3.
4.      for token in tokens:
5.
6.          kalimat = kalimat.replace(token, " "+token+" ")
7.          kalimat = re.sub(r"\s+", " ", kalimat)
8.          kalimat = re.sub(r"\s$", "", kalimat)
9.          simpan_file = open('IPOSTagger\kalimatinput.tag', 'w')
10.         simpan_file.write(kalimat)
11.         simpan_file.close()

```

Setelah penambahan spasi dilakukan pada baris dua sampai delapan, pada baris sembilan, hasil proses ini kemudian disimpan di file “kalimatinput.tag” di folder IPOSTagger untuk diproses menggunakan IPOSTagger.

Proses *tagging* PoS dilakukan pada *file* kalimat masukan yang disimpan dengan ekstensi *.tag pada proses sebelumnya. Proses ini menggunakan aplikasi IPOSTagger dengan korpus yang telah dilatih. Hasil keluaran *tagging* PoS disimpan pada *file* berekstensi *.hsl yang berisi “kata/PoS” dan ditampilkan di kolom kedua yang terletak dibawah kolom untuk memasukkan kalimat pada halaman *web* prediksi jeda. Proses dilakukan dengan memanggil perintah di *command prompt*. Kode program yang dipakai dapat dilihat dibawah ini.

Kode program 4.6 Proses *Tagging* PoS

```

1.  def prediksi_pos():
2.      print(os.getcwd())
3.      subprocess.call('dir', shell=True)
4.      curr_dir = os.chdir(os.getcwd()+"/IPOSTagger")
5.      subprocess.call('java ipostagger kalimatinput.tag 1 0 0 1 >
6.          '+hasil_tagger_POS, shell= True)

```

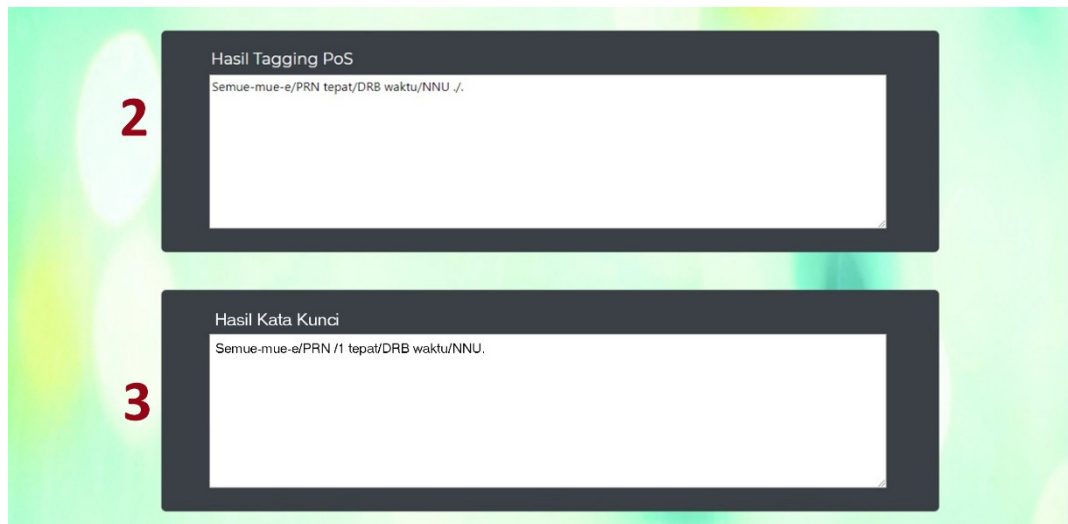
Proses IPOSTagger memanggil perintah “java ipostagger kalimatinput.tag 1 0 0 1” yang dipakai untuk melakukan fungsi pelabelan kata berdasarkan korpus yang telah dilatih. Hasil dari proses IPOSTagger ditampilkan di kolom kedua halaman web.

Gambar 4.11 Tampilan *output* proses *tagging* PoS

Pada gambar 4.61, hasil *tagging* PoS ditampilkan di kolom kedua. Hasil ini yang dipakai di proses kata kunci pada tahap selanjutnya.

4.1.6.2.2 Kata Kunci

Kalimat dimasukan kemudian diproses menggunakan aturan kata kunci. Pada tahap ini, aturan kata kunci diperiksa apakah ada kata yang memiliki jeda atau tidak sehingga hasil akhir proses berupa teks kalimat bahasa Melayu Pontianak dengan penandaan jeda singkat dan jeda panjang seperti pada gambar dibawah ini.



Gambar 4.12 Tampilan *output* hasil kata kunci

Hasil aturan kata kunci di kolom ketiga di proses untuk pemenggalan kata yang termasuk dalam aturan kata kunci dengan ditandai jeda singkat dan jeda panjang sehingga hasil akhir berupa teks kalimat bahasa Melayu serta langsung melakukan penghapusan label PoS.

4.1.6.2.3 Hasil Prediksi Jeda

Kalimat dimasukan kemudian diproses menggunakan aturan kata kunci. Pada tahap ini, aturan kata kunci diperiksa apakah ada kata yang memiliki jeda atau tidak sehingga hasil akhir proses berupa teks kalimat bahasa Melayu Pontianak dengan penandaan jeda singkat dan jeda panjang seperti pada gambar dibawah ini.



Gambar 4.13 Tampilan *output* hasil prediksi jeda

Hasil prediksi jeda pada kolom keempat di proses untuk penghapusan label PoS sehingga hasil akhir berupa teks kalimat bahasa Melayu Pontianak dengan penandaan jeda singkat dan jeda panjang.

4.1.6.2.4 Hasil Prediksi Jeda untuk Keperluan Sistem TTS

Pada tahap ini dilakukan normalisasi hasil prediksi jeda untuk mengubah tanda jeda singkat(/1) menjadi tanda “|” dan mengubah tanda jeda panjang (/2) menjadi tanda “||” untuk keperluan implementasi sistem *Text-to-Speech*.

Kode program 4.7 Hasil Prediksi Jeda

```

8.     with open('kal_prediksi_jeda.txt', 'r') as file :
9.         filedata = file.read()

10.        filedata = filedata.replace('/1', '|')
11.        filedata = filedata.replace('/2', '||')
12.        filedata = filedata.replace('/2', '||')

13.     with open('kal_prediksi_tts.txt', 'w+') as file:
14.         file.write(filedata)

```

Kode program 4.7 dipakai untuk normalisasi hasil jeda sehingga hasil akhir dari prediksi jeda berupa teks kalimat bahasa Melayu Pontianak dengan tanda baca | dan || untuk jeda. Hasil akhir dapat dilihat pada gambar dibawah ini.

Gambar 4.14 Hasil Prediksi untuk Keperluan TTS

Hasil prediksi jeda pada kolom kelima kemudian dinormalisasi menjadi hasil akhir prediksi jeda yang digunakan untuk keperluan implementasi sistem

TTS seperti pada kolom keempat.

4.2 Hasil Pengujian

4.2.1 Hasil Pengujian Kecocokan Frasa Jeda dalam Satu Kalimat Penuh

Pengujian kemunculan jeda dilakukan untuk melihat persamaan kemunculan kejadian jeda pada kalimat asli korpus yang sudah ditandai kejadian jeda sesuai ucapan penutur dan kalimat hasil prediksi dari aturan kata kunci. Total kalimat yang diuji sebanyak 500 kalimat dari kalimat penutur dan 500 kalimat hasil aturan kata kunci. Hasil perhitungan dimasukkan ke dalam persamaan 3.1 pada Bab 3 :

$$\text{Akurasi frasa jeda dalam kalimat} = \frac{\text{jumlah kalimat yang sesuai}}{\text{Total kalimat yang diuji}} \times 100 \%$$

Tabel pengujian yang dipakai mengacu pada tabel 3.7 pada Bab 3 untuk rencana pengujian pada frasa jeda ucapan dalam satu kalimat. Hasil nilai pengujian akurasi frasa jeda dalam kalimat dapat dilihat pada tabel berikut:

Tabel 4.14 Hasil Pengujian Jeda Penutur dan Jeda Kata Kunci Kecocokan Frasa dalam Satu Kalimat Penuh

Tipe Pengujian	Jumlah Kalimat	Akurasi (%)
Kemunculan jeda singkat dan panjang	500	41.6%
Kemunculan jeda panjang	500	82.2%

Dari hasil pengujian akurasi kecocokan jeda antara jeda dari penutur dan jeda dari informasi parafrasa didapatkan nilai akurasi dari total 500 kalimat dengan akurasi pada kemunculan jeda singkat dan panjang sebesar 41.6% dan kemunculan jeda panjang memiliki akurasi sebesar 82.2%. Hasil lengkap pengujian dapat dilihat pada Lampiran E.

Tabel 4.15 Hasil Pengujian Jeda Penutur dan Jeda Kata Kunci Kecocokan Frasa dalam Satu Kalimat Penuh Menggunakan PoS

Tipe Pengujian	Jumlah Kalimat	Akurasi (%)
Kemunculan jeda singkat dan panjang	500	54%

Kemunculan jeda panjang	500	80.2%
-------------------------	-----	-------

Dari hasil pengujian akurasi kecocokan jeda antara jeda dari penutur dan jeda dari informasi parafrasa didapatkan nilai akurasi dari total 500 kalimat dengan akurasi pada kemunculan jeda singkat dan panjang 54% dan kemunculan jeda panjang memiliki akurasi sebesar 80.2%. Hasil lengkap pengujian dapat dilihat pada Lampiran F.

Hasil pengujian akurasi kecocokan frasa jeda dengan *rule grammar* dan *rule* jeda yang baru dari penelitian sebelumnya oleh M. Iqbal Kamiluddin dapat dilihat pada tabel 4.16.

Tabel 4.16 Hasil Pengujian Jeda Penutur dengan Jeda Prediksi *Shallow Parsing*

Tipe Pengujian	Jumlah Kalimat	Akurasi (%)
Kemunculan jeda singkat dan panjang	500	10.6%

Dari hasil pengujian akurasi kecocokan jeda antara jeda dari penutur dan jeda dari *shallow parsing* dari penelitian sebelumnya didapatkan nilai akurasi dari total 500 kalimat dengan akurasi pada kemunculan jeda singkat dan panjang sebesar 10.6% atau sebanyak 53 kalimat. Hasil lengkap pengujian dapat dilihat pada Lampiran G.

4.2.1.1 Analisa Hasil Pengujian Kecocokan Frasa Jeda dalam Satu Kalimat

Penuh

Hasil pengujian menyatakan bahwa nilai akurasi terbaik pada tipe pengujian kemunculan jeda panjang. Tingginya nilai akurasi kemunculan jeda panjang dikarenakan frasa jeda panjang menghasilkan jeda pasti atau memiliki frasa yang sama pada penutur dari 500 kalimat. Sedangkan nilai pengujian kemunculan jeda singkat dan panjang menghasilkan nilai akurasi yang rendah, karena frasa jeda singkat menghasilkan pola ucapan berbeda pada setiap orang atau penutur memiliki jeda singkat yang bervariasi dari 500 kalimat sehingga aturan kata kunci dibuat belum dapat menangkap jeda singkat dengan baik.

Perbandingan hasil pengujian menyatakan bahwa nilai akurasi sebesar 10.6% pada pengujian jeda singkat dan jeda panjang. Pengujian jeda panjang tidak dilakukan karena pada penelitian sebelumnya, tidak dilakukan prediksi untuk jeda panjang, sehingga semua jeda dikategorikan sama. Rendahnya nilai akurasi sudah pernah dijelaskan di penelitian sebelumnya (Kamiludin, 2017) yang dikarenakan *rule grammar* yang dibuat berdasarkan pola struktur kalimat S-P-O-K. Hal ini yang menyebabkan *rule grammar* belum bisa membentuk frasa jeda yang sama dengan frasa jeda penutur.

4.2.2 Hasil Pengujian *Precision* , *Recall*, dan *F-Measure* Terhadap

Penggalan Frasa

Pengujian *recall* , *precision* , dan *F-Measure* dilakukan dengan membandingkan korpus frasa jeda penutur dan korpus hasil prediksi sistem. Pengujian dilakukan sebanyak 2 kali yakni pada frasa jeda singkat (1) dan frasa jeda panjang(2). Pada pengujian terhadap 5 tahap yaitu membandingkan korpus uji sebanyak 100 kalimat, 200 kalimat, 400 kalimat , dan 500 kalimat.

4.2.2.1 Hasil Pengujian Penggalan Frasa Kata Kunci Terhadap Jeda Singkat Dan Panjang

Pengujian pemenggalan frasa tanpa menggunakan PoS terhadap jeda singkat dan panjang dilakukan sebanyak 5 kali yaitu, membandingkan kalimat korpus sebanyak 100 kalimat, 200 kalimat, 300 kalimat, 400 kalimat, dan 500

kalimat. Hasil pengujian keseluruhan bisa dilihat pada lampiran H. Tabel hasil pengujian terhadap 100 kalimat dari 500 korpus analisis dapat dilihat pada tabel 4.16 sebagai berikut:

Tabel 4.17 Pengujian Terhadap 100 Kalimat dari 500 Kalimat Korpus Analisis

No.	Kalimat Korpus Uji	Sesuai		Tidak Sesuai	Precision ($\frac{a}{a+c}$)	Recall ($\frac{a}{a+b}$)	F-Measure ($2 \cdot \frac{p \cdot r}{p+r}$)
		Terambil (a)	Tidak Terambil (b)	Terambil (c)			
1.	0-100	183	96	97	0.6535	0.6559	0.6547
2.	101-200	188	146	134	0.5838	0.5628	0.5731
3.	201-300	188	95	95	0.6643	0.6643	0.6643
4.	301-400	167	112	119	0.5839	0.5985	0.5911
5.	401-500	146	102	120	0.5488	0.5863	0.5669
Rata-rata					0.60686	0.61356	0.6100

Pengujian dilakukan dengan menghitung *precision*, *recall*, dan *f-measure* terhadap 100 kalimat dari 500 kalimat, kemudian dari 500 kalimat dibagi menjadi 5 bagian yaitu 0-100 kalimat, 101-200 kalimat, 201-300 kalimat, 301-400 kalimat, dan 401-500 kalimat, dan didapatkan nilai rangkuman pengujian potongan frasa kata kunci terhadap jeda singkat. Tabel hasil rangkuman nilai pengujian potongan frasa kata kunci tanpa PoS terhadap jeda singkat dapat dilihat pada tabel 4.17 sebagai berikut:

Tabel 4.18 Rangkuman Nilai Pengujian Potongan Frasa Kata Kunci Terhadap Jeda Singkat dan Panjang

No.	Kalimat Korpus Uji	Precision ($\frac{a}{a+c}$)	Recall ($\frac{a}{a+b}$)	F-Measure ($2 \cdot \frac{p \cdot r}{p+r}$)
1.	100	0.6068	0.6135	0.6100
2.	200	0.5429	0.6097	0.6066
3.	300	0.6026	0.6083	0.6054
4.	400	0.6026	0.6082	0.6053
5.	500	0.6026	0.6081	0.6053

Dari hasil pengujian *Precision*, *Recall*, dan *F-Measure* pemenggalan frasa terhadap jeda 1 dari 100, 200, 300, 400 dan 500 kalimat didapatkan nilai *F-Measure* pada 100 kalimat sebesar 0.6100 , pada 200 kalimat sebesar 0.6066, pada 300 kalimat sebesar 0.6654, pada 400 kalimat sebesar 0.6053, dan pada 500 kalimat sebesar 0.6053.

4.2.2.2 Hasil Pengujian Penggalan Frasa Kata Kunci Terhadap Jeda Singkat dan Jeda Panjang Menggunakan PoS

Pengujian pemenggalan frasa menggunakan PoS terhadap jeda singkat dilakukan sebanyak 5 kali yaitu, membandingkan kalimat korpus sebanyak 100 kalimat, 200 kalimat, 300 kalimat, 400 kalimat, dan 500 kalimat. Hasil pengujian keseuruhan bisa dilihat pada lampiran I.

Tabel hasil pengujian terhadap 100 kalimat dari 500 korpus analisis dapat dilihat pada tabel 4.18 sebagai berikut:

Tabel 4.19 Pengujian Terhadap 100 Kalimat dari 500 Kalimat Korpus Analisis

No	Kalimat Korpus Uji	Sesuai		Tidak Sesuai	Precision $(\frac{a}{a+c})$	Recall $(\frac{a}{a+b})$	F-Measure $(2 \cdot \frac{p \cdot r}{p+r})$
		Terambil (a)	Tidak Terambil (b)	Terambil(c)			
1.	0-100	184	95	95	0.6501	0.6594	0.6548
2.	101-200	190	144	135	0.5846	0.5688	0.5766
3.	201-300	188	95	95	0.6643	0.6643	0.6643
4.	301-400	165	114	124	0.5709	0.5913	0.5809
5.	401-500	146	103	122	0.5447	0.5863	0.5647
Rata-rata					0.6029	0.6140	0.6082

Pengujian dilakukan dengan menghitung *precision*, *recall*, dan *f-measure* terhadap 100 kalimat dari 500 kalimat, kemudian dari 500 kalimat dibagi menjadi 5 bagian yaitu 0-100 kalimat, 10-200 kalimat, 201-300 kalimat, 301-400 kalimat, dan 401-500 kalimat, dan didapatlah nilai rangkuman pengujian potongan frasa kata kunci terhadap jeda singkat menggunakan PoS. Tabel hasil rangkuman nilai pengujian potongan frasa kata kunci terhadap jeda singkat menggunakan PoS

dapat dilihat pada tabel 4.19 sebagai berikut:

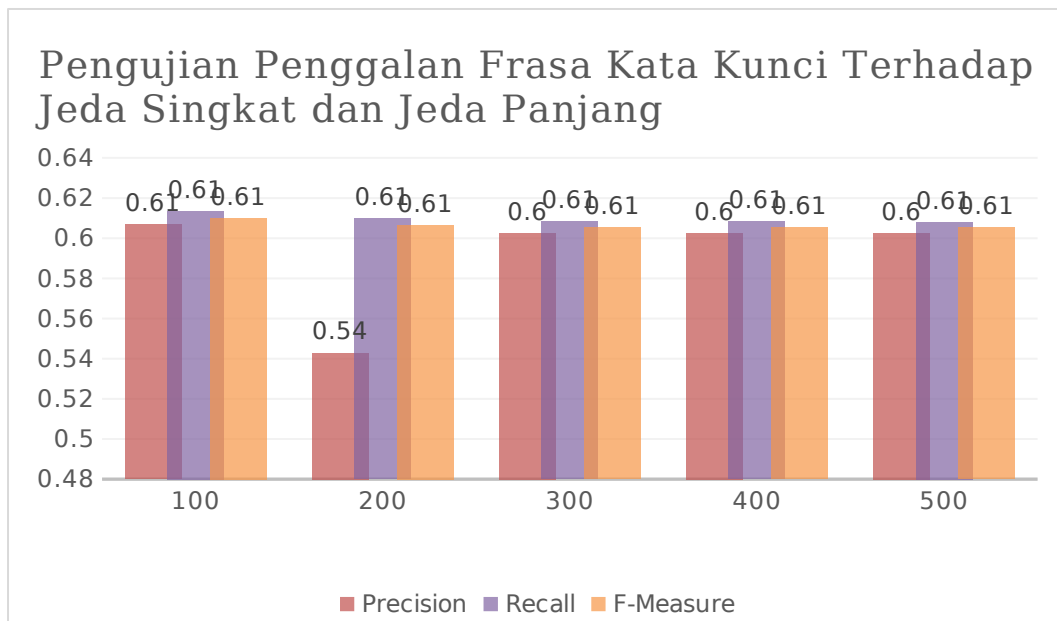
Tabel 4.20 Rangkuman Nilai Pengujian Potongan Frasa Kata Kunci Terhadap Jeda Singkat Menggunakan PoS

No.	Kalimat Korpus Uji	Precision ($\frac{a}{a+c}$)	Recall ($\frac{a}{a+b}$)	F-Measure ($2 \cdot \frac{p \cdot r}{p+r}$)
1.	100	0.6029	0.6140	0.6082
2.	200	0.5998	0.5492	0.6049
3.	300	0.5811	0.5949	0.5895
4.	400	0.5683	0.5732	0.5709
5.	500	0.5987	0.6088	0.6037

Dari hasil pengujian *Precision*, *Recall*, dan *F-Measure* pemenggalan frasa terhadap jeda 1 dari 100, 200, 300, 400 dan 500 kalimat didapatkan nilai *F-Measure* pada 100 kalimat sebesar 0.6082 , pada 200 kalimat sebesar 0.6049, pada 300 kalimat sebesar 0.5895, pada 400 kalimat sebesar 0.5709, dan pada 500 kalimat sebesar 0.6037.

4.2.2.1.1 Analisis Hasil Pengujian Penggalan Frasa Kata Kunci Terhadap Jeda Singkat dan Panjang

Hasil pengujian menyatakan bahwa nilai akurasi memiliki rata-rata yang hampir sama untuk pengujian 100-500 kalimat dengan kisaran nilai 0.60. Tampilan hasil pengujian penggalan frasa terhadap jeda singkat dapat dilihat pada gambar grafik berikut.

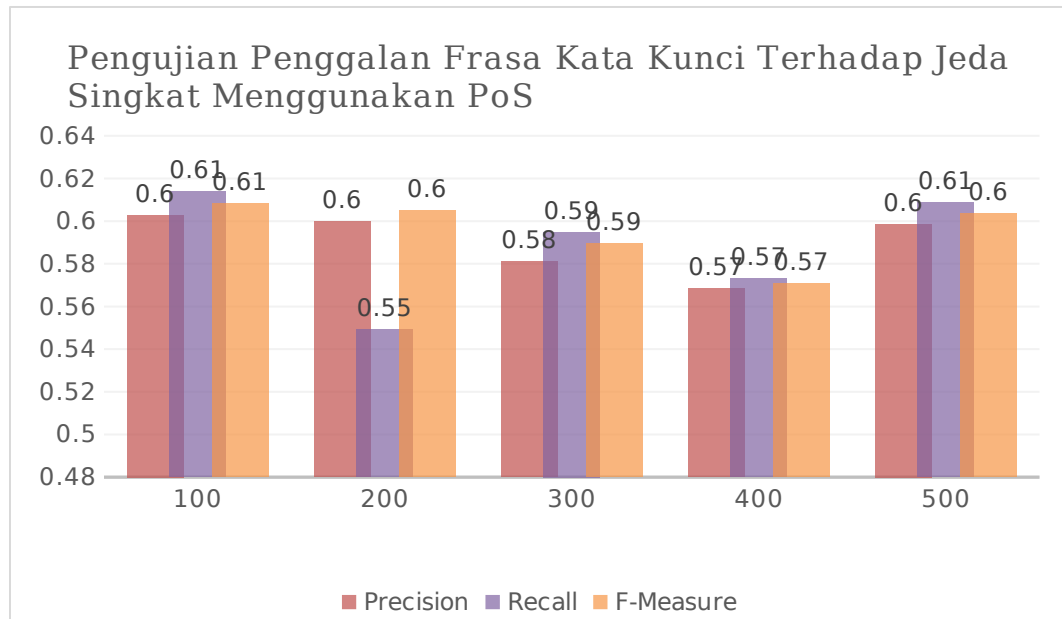


Gambar 4.15 Grafik Hasil Pengujian Penggalan Frasa Kata Kunci Terhadap Jeda Singkat dan Jeda Panjang

Nilai pada pengujian yang berada dikisaran 0.60 disebabkan jeda singkat memiliki jeda yang tak pasti atau selalu berubah pada setiap pengucapan sehingga banyak kalimat yang tidak akurat. Frasa jeda yang salah ini dikarenakan aturan kata kunci membentuk frasa sesuai banyaknya kemunculan kata yang memiliki jeda yang berbeda pada penutur. Seperti yang telah dijelaskan di poin 4.2.1.1 jeda singkat menghasilkan pola ucapan berbeda pada setiap orang atau penutur atau memiliki jeda singkat yang bervariasi dari 500 kalimat sehingga aturan kata kunci dibuat belum dapat menangkap jeda singkat dengan baik.

4.2.2.1.2 Analisis Hasil Pengujian Penggalan Frasa Kata Kunci Terhadap Jeda Singkat Menggunakan PoS

Hasil pengujian menyatakan bahwa nilai akurasi memiliki rata-rata yang hampir sama untuk pengujian 100-500 kalimat dengan kisaran nilai 0.59. Tampilan hasil pengujian penggalan frasa terhadap jeda singkat dapat dilihat pada gambar grafik berikut.



Gambar 4.16 Grafik Hasil Pengujian Penggalan Frasa Kata Kunci Terhadap Jeda Singkat Menggunakan PoS

Nilai pada pengujian yang berada dikisaran 0.59 disebabkan jeda singkat memiliki jeda yang tak pasti atau selalu berubah pada setiap pengucapan sehingga banyak kalimat yang tidak akurat. Frasa jeda yang salah ini dikarenakan aturan kata kunci membentuk frasa sesuai banyaknya kemunculan kata yang memiliki jeda yang berbeda pada penutur. Seperti yang telah dijelaskan di poin 4.2.1.1 jeda singkat menghasilkan pola ucapan berbeda pada setiap orang atau penutur atau memiliki jeda singkat yang bervariasi dari 500 kalimat sehingga aturan kata kunci dibuat belum dapat menangkap jeda singkat dengan baik.

4.2.2.3 Hasil Pengujian Penggalan Frasa Kata Kunci Terhadap Jeda Panjang

Pengujian pemenggalan frasa tanpa menggunakan PoS terhadap jeda 2 dilakukan sebanyak 5 kali yaitu, membandingkan kalimat korpus sebanyak 100 kalimat, 200 kalimat, 300 kalimat, 400 kalimat, dan 500 kalimat. Hasil pengujian keseluruhan bisa dilihat pada lampiran J. Tabel hasil pengujian terhadap 100 kalimat dari 500 korpus analisis dapat dilihat pada tabel 4.20 sebagai berikut:

Tabel 4.21 Hasil Pengujian terhadap 100 kalimat dari 500 kalimat korpus

No	Kalimat Korpus Uji	Sesuai		Tidak Sesuai Terambil(c)	Precision $(\frac{a}{a+c})$	Recall $(\frac{a}{a+b})$	F-Measure $(2. \frac{p.r}{p+r})$
		Terambil (a)	Tidak Terambil (b)				
1.	0-100	153	24	13	0.922	0.864	0.892
2.	101-200	140	49	25	0.848	0.741	0.791
3.	201-300	141	39	21	0.870	0.783	0.824
4.	301-400	156	49	29	0.843	0.761	0.800
5.	401-500	129	42	25	0.837	0.754	0.794
Rata-rata					0.864	0.782	0.821

Pengujian dilakukan dengan menghitung *precision*, *recall*, dan *f-measure* terhadap 100 kalimat dari 500 kalimat, kemudian dari 500 kalimat dibagi menjadi 5 bagian yaitu 0-100 kalimat, 10-200 kalimat, 201-300 kalimat, 301-400 kalimat, dan 401-500 kalimat, dan didapatkan nilai rangkuman pengujian potongan frasa kata kunci terhadap jeda panjang. Tabel hasil rangkuman nilai pengujian potongan frasa kata kunci menggunakan PoS terhadap jeda panjang kata kunci dapat dilihat pada tabel 4.21 sebagai berikut:

Tabel 4.22 Rangkuman Nilai Pengujian Pemenggalan Frasa Kata Kunci Terhadap Jeda Panjang

No.	Kalimat Korpus Uji	Precision $(\frac{a}{a+c})$	Recall $(\frac{a}{a+b})$	F-Measure $(2. \frac{p.r}{p+r})$
1.	100	0.864	0.782	0.821
2.	200	0.864	0.781	0.820
3.	300	0.864	0.781	0.820
4.	400	0.864	0.780	0.819

No.	Kalimat Korpus Uji	Precision $(\frac{a}{a+c})$	Recall $(\frac{a}{a+b})$	F-Measure $(2. \frac{p.r}{p+r})$
5.	500	0.864	0.781	0.820

Dari hasil pengujian *Precision*, *Recall*, dan *F-Measure* pemenggalan frasa terhadap jeda 1 dari 100, 200, 300, 400 dan 500 kalimat didapatkan nilai *F-Measure* pada 100 kalimat sebesar 0.821 , pada 200 kalimat sebesar 0.820, pada 300 kalimat sebesar 0.820, pada 400 kalimat sebesar 0.819, dan pada 500 kalimat sebesar 0.820.

4.2.2.4 Hasil Pengujian Penggalan Frasa Kata Kunci Terhadap Jeda Panjang Menggunakan PoS

Pengujian pemenggalan frasa kata kunci terhadap jeda panjang menggunakan PoS dilakukan sebanyak 5 kali yaitu, membandingkan kalimat korpus sebanyak 100 kalimat, 200 kalimat, 300 kalimat, 400 kalimat, dan 500 kalimat. Hasil pengujian keseluruhan bisa dilihat pada lampiran K. Tabel rangkuman hasil pengujian dapat dilihat pada tabel 4.22 sebagai berikut:

Tabel 4.23 Pengujian Terhadap 100 Kalimat dari 500 Kalimat Korpus Analisis

No	Kalimat Korpus Uji	Sesuai		Tidak Sesuai Terambil(c)	Precision $(\frac{a}{a+c})$	Recall $(\frac{a}{a+b})$	F-Measure $(2. \frac{p.r}{p+r})$
		Terambil (a)	Tidak Terambil (b)				
1.	0-100	152	25	18	0.8941	0.8587	0.8760
2.	101-200	138	51	29	0.8263	0.7301	0.7752
3.	201-300	140	40	24	0.8536	0.7777	0.8139
4.	301-400	156	49	29	0.8432	0.7609	0.8
5.	401-500	131	40	25	0.8397	0.7660	0.8012
Rata-rata					0.8513	0.7786	0.8136

Pengujian dilakukan dengan menghitung *precision*, *recall*, dan *f-measure* terhadap 100 kalimat dari 500 kalimat, kemudian dari 500 kalimat dibagi menjadi 5 bagian yaitu 0-100 kalimat, 10-200 kalimat, 201-300 kalimat, 301-400 kalimat,

dan 401-500 kalimat, dan didapatkan nilai rangkuman pengujian potongan frasa kata kunci terhadap jeda panjang menggunakan PoS. Tabel hasil rangkuman nilai pengujian potongan frasa kata kunci terhadap jeda panjang menggunakan PoS dapat dilihat pada tabel 4.23 sebagai berikut:

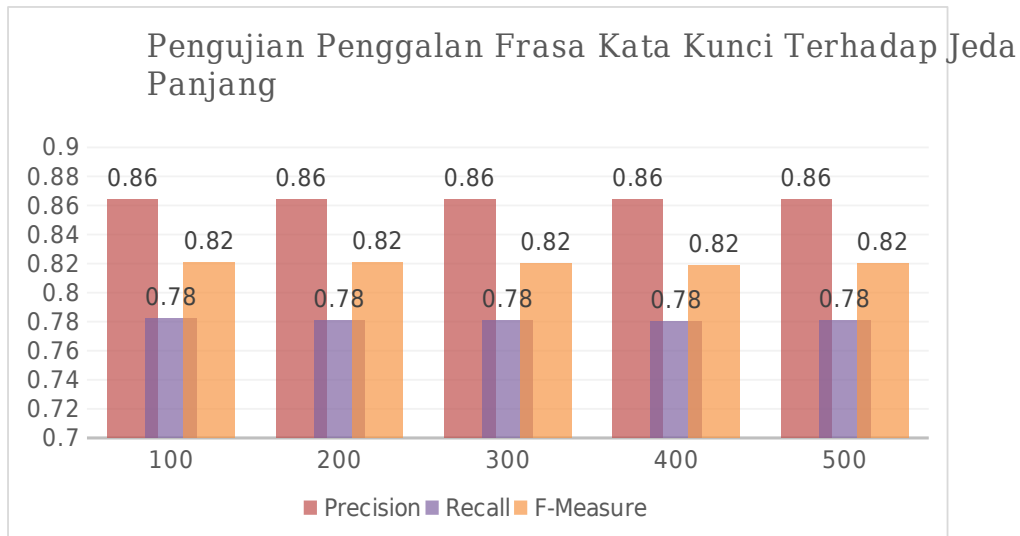
Tabel 4.24 Rangkuman Nilai Pengujian Potongan Frasa Kata Kunci Terhadap Jeda Panjang Menggunakan PoS

No.	Kalimat Korpus Uji	Precision ($\frac{a}{a+c}$)	Recall ($\frac{a}{a+b}$)	F-Measure ($2 \cdot \frac{p \cdot r}{p+r}$)
1.	100	0.8513	0.7786	0.8136
2.	200	0.8480	0.7749	0.8090
3.	300	0.8217	0.7498	0.7836
4.	400	0.8467	0.7734	0.8084
5.	500	0.8467	0.7733	0.8083

Dari hasil pengujian *Precision*, *Recall*, dan *F-Measure* pemenggalan frasa terhadap jeda 1 dari 100, 200, 300, 400 dan 500 kalimat didapatkan nilai *F-Measure* pada 100 kalimat sebesar 0.8136 , pada 200 kalimat sebesar 0.8090, pada 300 kalimat sebesar 0.7836, pada 400 kalimat sebesar 0.8084, dan pada 500 kalimat sebesar 0.8083.

4.2.2.2.1 Analisis Hasil Pengujian Penggalan Frasa Kata Kunci Terhadap Jeda Panjang

Hasil pengujian menyatakan bahwa nilai akurasi memiliki rata-rata yang hampir sama dengan kisaran nilai 0.82. Tampilan hasil pengujian penggalan frasa terhadap jeda panjang dapat dilihat pada gambar grafik berikut.

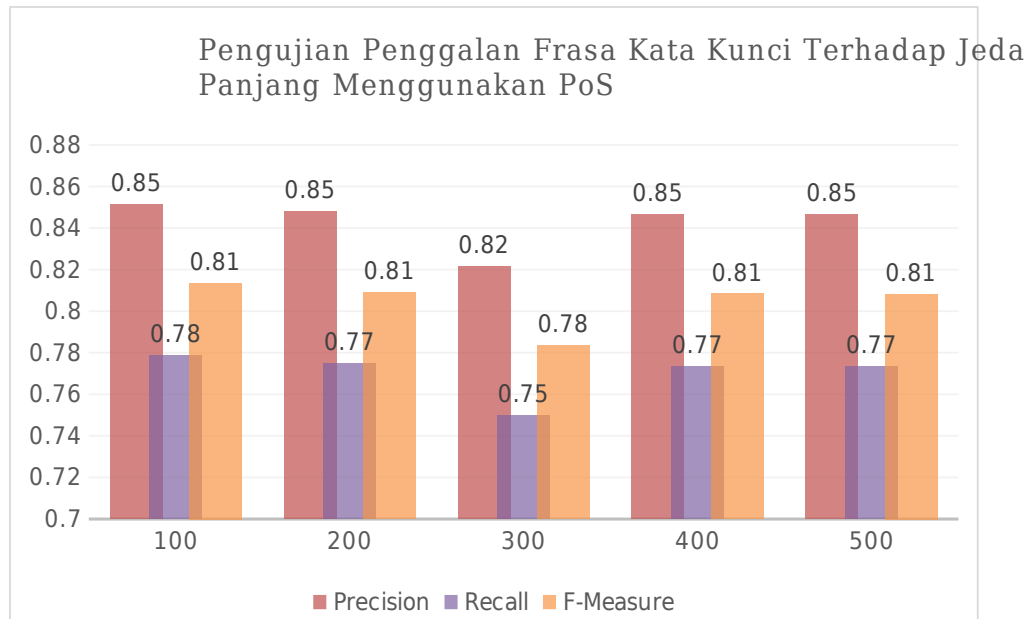


Gambar 4.17 Grafik Hasil Pengujian Hasil Frasa Kata Kunci Terhadap Jeda Panjang

Nilai pengujian *precision* menunjukkan angka yang sama di nilai 0.86 yang berarti aturan kata kunci berhasil memprediksi penggalan frasa yang tepat untuk semua kalimat baik dari 100 kalimat sampai 500 kalimat. Nilai *recall* lebih rendah, karena masih ada penggalan frasa yang tidak sesuai dengan frasa jeda penutur karena aturan yang belum cocok. Nilai *f-measure* memiliki nilai yang hampir sama dan tergolong baik yakni 0.82. Prediksi jeda panjang memiliki nilai yang lebih baik dan tinggi, karena berdasarkan jeda penutur, letak jeda panjang cenderung memiliki pola berhenti di suatu frasa yang sama, sehingga aturan kata kunci dan aturan jeda yang dibuat dapat memprediksi jeda dengan baik.

4.2.2.2.2 Analisis Hasil Pengujian Penggalan Frasa Terhadap Jeda Panjang Menggunakan PoS

Hasil pengujian menyatakan bahwa nilai akurasi memiliki rata-rata yang hampir sama dengan kisaran nilai 0.80. Tampilan hasil pengujian penggalan frasa terhadap jeda panjang dapat dilihat pada gambar grafik berikut.



Gambar 4.18 Grafik Hasil Pengujian Penggalan Frasa Kata Kunci Terhadap Jeda Panjang Menggunakan PoS

Nilai pengujian *precision* menunjukkan angka yang sama di nilai 0.84 yang berarti aturan kata kunci berhasil memprediksi penggalan frasa yang tepat untuk semua kalimat baik dari 100 kalimat sampai 500 kalimat. Nilai *recall* lebih bervariasi, karena masih ada penggalan frasa yang tidak sesuai dengan frasa jeda penutup karena aturan yang belum cocok. Nilai *f-measure* memiliki nilai yang hampir sama dan tergolong baik yakni 0.80. Prediksi jeda panjang memiliki nilai yang lebih baik dan tinggi, karena berdasarkan jeda penutup, letak jeda panjang cenderung memiliki pola berhenti di suatu frasa yang sama, sehingga aturan kata kunci yang dibuat dapat memprediksi jeda dengan baik.

4.3 Analisa Hasil Keseluruhan Pengujian

1. Berdasarkan pengujian kecocokan frasa pada satu kalimat, nilai akurasi prediksi jeda kata kunci dan kata kunci PoS untuk jeda singkat dan panjang berbeda. Perbandingan nilai akurasi dapat dilihat pada tabel 4.25 sebagai berikut:

Tabel 4.25 Perbandingan Nilai Akurasi Kata Kunci dan Kata Kunci PoS

	Kata Kunci	Kata Kunci PoS
Frasa Jeda Singkat dan Panjang	41.6% atau sebanyak 208 kalimat	54% atau sebanyak 270 kalimat
Frasa Jeda Panjang	82.2 % atau sebanyak 411 kalimat	80.2% atau sebanyak 401 kalimat

Berdasarkan pada tabel 4.25, penggunaan kata kunci PoS lebih baik dikarenakan penggunaan kata kunci PoS melakukan pemenggalan frasa lebih signifikan dibanding hanya kata kunci saja. Perbedaannya adalah ada satu kata yang memiliki 2 PoS yang berbeda contoh kalimatnya:

- a. Name/NNC aku/PRP bekincih/VBI macam/RP **orang/NNC** kenak/VBT masok-ek/VBT antu/NNC.
- b. Padahal/RP yang/CON betandeng/VBI tu/DT cume/DRB sebelas/CDP **orang/CDA** lawan/VBT sebelas/CDP orang/CDA jak/RP ./.

Penggunaan kata “orang” disini berbeda karena menyesuaikan kalimat dan kata sebelum dan sesudahnya. Pada kalimat a kata orang diartikan sebagai seseorang yang mengalami kejadian tersebut, sedangkan b diartikan sebagai sekumpulan orang yang mengalami kejadian tersebut. Oleh karena itu, ada perbedaan hasil antara penggunaan PoS dan tidak.

2. Berdasarkan pengujian *precision*, *recall*, dan *f-measure*, dari kemunculan jeda dan penggalan frasa jeda, analisis yg dapat disimpulkan yaitu nilai akurasi prediksi jeda kata kunci dan kata kunci PoS untuk jeda singkat dan jeda panjang berbeda. Perbandingan nilai akurasi dapat dilihat pada tabel 4.26 sebagai berikut:

Tabel 4.26 Perbandingan nilai akurasi jeda singkat dan jeda panjang

	Kata Kunci			Kata Kunci PoS		
	Precision	Recall	F-Measure	Precision	Recall	F-Measure
Frasa Jeda Singkat dan Panjang	0.58	0.59	0.60	0.63	0.61	0.59
Frasa Jeda Panjang	0.86	0.78	0.82	0.84	0.77	0.80

Berdasarkan pada tabel 4.26, nilai akurasi terbaik dihasilkan pada prediksi jeda panjang. Hal ini dikarenakan, jeda yang terjadi dapat berupa tanda baca atau banyak kata yang memberikan penekanan dikata tersebut pada setiap ucapan dan jeda yang terbentuk bisa mendapatkan nilai yang sama dengan penutur. Akurasi yang rendah pada jeda singkat dikarenakan penentuan aturan kata kunci yang belum sempurna pada kalimat sehingga belum bisa memotong frasa jeda secara akurat sesuai frasa jeda penutur.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil analisis dan pengujian terhadap prediksi jeda ucapan kalimat Bahasa Melayu Pontianak dengan menggunakan pemenggalan informasi parafrasa, maka dapat ditarik kesimpulan sebagai berikut:

1. Aturan kata kunci dan aturan kata kunci PoS dapat menghasilkan frasa jeda pada kalimat bahasa Melayu Pontianak sesuai dengan aturan kata kunci yang disiapkan.
2. Berdasarkan analisis yang telah dilakukan untuk aturan kata kunci terdapat 217 kata yang dianalisis dan 122 yang masuk dalam aturan kata kunci tanpa PoS, sedangkan aturan kata kunci menggunakan PoS terdapat 224 kata yang dianalisis dan 126 yang masuk dalam aturan kata kunci.
3. Berdasarkan pengujian kecocokan frasa pada satu kalimat, hasil pengujian akurasi kecocokan jeda terhadap 500 kalimat menggunakan kata kunci sebesar 41.6% atau sebanyak 208 kalimat untuk uji kemunculan jeda singkat dan panjang, dan nilai akurasi sebesar 82.2% atau sebanyak 411 kalimat untuk uji kemunculan jeda panjang. Sedangkan hasil kecocokan akurasi kecocokan jeda dalam satu kalimat terhadap 500 kalimat menggunakan kata kunci PoS sebesar 54% atau sebanyak 270 kalimat untuk uji kemunculan jeda singkat dan panjang, dan nilai akurasi sebesar 80.2% atau sebanyak 401 kalimat untuk uji kemunculan jeda panjang. Hasil lengkap pengujian dapat dilihat pada lampiran E dan lampiran F.
4. Berdasarkan penelitian yang telah dilakukan, set PoS yang banyak terjadi kejadian jeda terdapat pada set PoS RP dan VBT yaitu sebanyak 15 dan 16 kata, hal itu dikarenakan karena pada ucapan penutur banyak memberikan penekanan pada kata yang termasuk set PoS tersebut untuk memperjelas kalimat yang ingin disampaikan.
5. Prediksi jeda pada ucapan dengan menggunakan informasi parfarasa memiliki nilai akurasi yang baik yaitu nilai akurasi jeda untuk jeda

panjang sebesar 82.2%, dan jeda panjang yang memiliki nilai *precision*, *recall*, dan *f-measure* yang cukup tinggi yaitu sebesar 0.86, 0.78 dan nilai *f-measure* sebesar 0.82.

6. Nilai *precision*, *recall*, dan *F-measure* belum mampu menghasilkan nilai akurasi yang baik dikarenakan aturan kata kunci yang dirancang masih belum sempurna untuk menghasilkan potongan frasa jeda. Hal ini disebabkan penentuan aturan kata kunci yang belum sempurna pada kalimat sehingga belum bisa memotong frasa jeda secara akurat sesuai frasa jeda penutur.

5.2 Saran

Adapun hal yang menjadi saran dalam pengembangan sistem ini agar lebih baik lagi sebagai berikut:

1. Perlu pengembangan sistem *text to speech* bahasa melayu pontianak dengan memanfaatkan informasi berupa jeda yang dihasilkan oleh aturan kata kunci dalam penelitian ini agar suara yang dihasilkan oleh sistem TTS lebih mudah dipahami.
2. Perlu adanya perbandingan dengan metode lain dalam pembentukan jeda sehingga dapat dilihat mana yang lebih baik dari kecepatan proses maupun kemunculan pengguna metode.
3. Perlu adanya pengembangan PoS yang lebih spesifik untuk pelabelan kelas kata dalam kalimat sehingga jeda yang dihasilkan lebih akurat dan dapat menyesuaikan jeda penutur untuk prediksi jeda singkat dan jeda panjang sehingga dapat diimplementasikan untuk memprediksi bahasa lain selain Bahasa Melayu Pontianak.