**PERBANDINGAN METODE POS TAGGER TERHADAP NILAI AKURASI UNTUK BAHASA INDONESIA**

**SKRIPSI**

Jurusan Informatika

Program Studi Sarjana Informatika

Oleh:

**NUR RAHMAN R**

NIM D1042131018



**FAKULTAS TEKNIK**

**UNIVERSITAS TANJUNGPURA**

**PONTIANAK**

**2020**

Halaman Pernyataan

Yang bertanda tangan di bawah ini:

Nama : Nur Rahman R

NIM : D1042131018

Menyatakan bahwa dalam skripsi yang berjudul “ Perbandingan Metode PoS Tagger Terhadap Nilai Akurasi Untuk Bahasa Indonesia” tidak terdapat karya yang pernah diajukan untuk memperoleh gelar sarjana di suatu perguruan tinggi manapun. Sepanjang pengetahuan Saya, tidak terdapat karya atau pendapat yang pernah ditulis atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam Daftar Pustaka.

Demikian pernyataan ini dibuat dengan sebenar-benarnya. Saya sanggup menerima konsekuensi akademis dan hukum di kemudian hari apabila pernyataan yang dibuat ini tidak benar.

Pontianak, Desember 2020

NUR RAHMAN R

NIM D1042131018

KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN

UNIVERSITAS TANJUNGPURA

**FAKULTAS TEKNIK**

Jalan Prof. Dr. H. Hadari Nawawi Pontianak 78124

Telepon (0561) 740186 Faximili (0561) 740186

Email : [ft@untan.ac.id](mailto:ft@untan.ac.id) Website : [http://teknik.untan.ac.id](http://teknik.untan.ac.id/)



**HALAMAN PENGESAHAN**

“PERBANDINGAN METODE POS TAGGER TERHADAP NILAI AKURASI UNTUK BAHASA INDONESIA”

Jurusan Informatika

Program Studi Sarjana Informatika

Oleh : “Nur Rahman R”

NIM. D1042131018

Telah dipertahankan didepan Penguji Skripsi pada tanggal 28 Desember 2020 dalam sidang secara daring (*online*) dan diterima sebagai salah satu persyaratan untuk memperoleh gelar sarjana.

Susunan Penguji Skripsi :

|  |  |  |
| --- | --- | --- |
| Dosen Pembimbing Utama | : Dr. Arif Bijaksana Putra Negara, S.T., M.T., | (NIP.197208081998021002) |
| Dosen Pembimbing Kedua | : Hafiz Muhardi, ST, M.Kom | (NIDK.888537001) |
| Dosen Penguji Utama | : Helfi Nasution, S.Kom. M.Cs. | (NIP.197104291998021002) |
| Dosen Penguji Kedua | : Yulianti, S.Kom., MMSI | (NIP.197210162008012005) |
|  |  |  |

Wakil Dekan Bidang Akademik,

Dr.-Ing. Ir. Slamet Widodo, M.T.,IPM.  
NIP. 196712231992031002

Pontianak, Desember 2020

Dekan,

Dr.rer.nat. Ir. R. M. Rustamaji, M.T., IPU

NIP. 196801161994031003

**HALAMAN PERSEMBAHAN**

Skripsi ini saya persembahkan untuk kedua orang tua saya, abang saya dan pacar saya yang tiada hentinya memberikan dukungan selama menjalani kuliah.

Terima kasih saya ucapkan kepada rekan-rekan mahasiswa Informatika Universitas Tanjungpura angkatan 2013 yang telah saling mendukung selama perkuliahan, saat skripsi, dan di luar bidang akademik.

Serta seluruh kerabat, dosen, staff TU, dan pihak yang telah membantu, yang tidak dapat disebutkan satu per satu. Semoga Tuhan senantiasa memberikan balasan kebaikan.

Kata Pengantar

Puji dan syukur penulis panjatkan kepada Tuhan Yang Maha Esa, karena berkat rahmat-Nya, penelitian dan penulisan skripsi yang berjudul “Perbandingan Metode PoS Tagger Terhadap Nilai Akurasi Untuk Bahasa Indonesia” ini dapat diselesaikan sebagai syarat untuk mengambil gelar Sarjana Komputer (S.Kom.).

Melalui penelitian ini, ucapan terimakasih saya sampaikan kepada pihak-pihak yang telah membantu saya baik berupa materil maupun dukungan moril yaitu kepada Bapak **Dr. Arif Bijaksana Putra Negara, S.T., M.T.,** selaku Dosen Pembimbing Pertama saya, yang telah memberikan banyak bimbingan, saran dan masukan yang membangun dalam melakukan perancangan dan pembuatan skripsi ini, dan kepada Bapak **Hafiz Muhardi, ST, M.Kom** selaku Dosen Pembimbing Kedua sekaligus yang telah memberikan banyak bimbingan, membantu dalam menyelesaikan penelitian ini.

Kepada Dosen Penguji Pertama Bapak **Helfi Nasution, S.Kom. M.Cs.** dan Dosen Penguji Pendamping Ibu **Yulianti, S.Kom., MMSI** yang telah memberikan saran serta masukan untuk perbaikan skripsi ini. Tentunya saya sadari terdapat banyak kekurangan dalam penyusunan skripsi ini, oleh karena itu kritik dan saran saya butuhkan. Semoga skripsi ini bermanfaat bagi perkembangan ilmu pengetahuan.

Pontianak, Desember 2020

Penulis,

NUR RAHMAN R

NIM. D1042131018

Abstrak­­

Metode PoS Tagger adalah metode yang digunakan sebagai acuan untuk melakukan tagging pos secara langsung sesuai dengan data uji yang dan data latih yang telah dilakukan pengujian untuk mendeteksi seberapa akurat metode pos tagger tersebut dalam melakukan tagging, yang dimana hal ini lakukan pengecekan dengan data tagging secara manual untuk melihat keakuratan dari metode tersebut. Data kalimat teks korpus yang digunakan sebanyak 500 kalimat bahasa Indonesia yang dimana kalimat teks digunakan dalam penelitian ini sebagai kalimat latih maupun kalimat uji. Penelitian ini melakukan pengembangan korpus yang berjumlah 500 kalimat, dan pengembangan tag PoS yang menggunakan data tag PoS dari penelitian lainnya kemudian hasil dari analisa penelitian ini adalah mendapatkan nilai akurasi dari setiap metode PoS tagger.

Kata Kunci : Pos Tagger, Bahasa Indonesia, Metode PoS Tagger, Nilai Akurasi.

Abstract­­

The PoS Tagger method is a method used as a reference for carrying out posttagging directly according to the test data and training data that has been tested to detect how accurate the tagger post method is in tagging, which is to check with tagging data manually to see the accuracy of the method. The data on the corpus text sentence data used are 500 Indonesian sentences, where the text sentences were used in this study as training sentences and test sentences. This study developed a corpus from 500 sentences, and developed a PoS tag using PoS tag data from other studies. Then the results of the analysis of this study were to obtain the accuracy value of each PoS tagger method.

Keywords : PoS Tagger, Indonesian, PoS Tagger Method, Accuracy Value.

Daftar Isi

[Halaman Pernyataan ii](#_Toc63963070)

[Halaman Pengesahan iii](#_Toc63963071)

[Halaman Persembahan iv](#_Toc63963072)

[Kata Pengantar v](#_Toc63963073)

[Abstrak vi](#_Toc63963074)

[Abstract vii](#_Toc63963075)

[Daftar Isi viii](#_Toc63963076)

[Daftar Tabel x](#_Toc63963077)

[Daftar Gambar xi](#_Toc63963078)

[Daftar Kode Program xii](#_Toc63963079)

[Daftar Grafik xiii](#_Toc63963080)

[BAB I Pendahuluan 1](#_Toc63963081)

[1.1 Latar Belakang 1](#_Toc63963082)

[1.2 Rumusan Masalah 3](#_Toc63963083)

[1.3 Tujuan Penelitian 3](#_Toc63963084)

[1.4 Pembatasan Masalah 3](#_Toc63963085)

[1.5 Sistematika Penulisan 3](#_Toc63963086)

[BAB II Tinjauan Pustaka 5](#_Toc63963087)

[2.1 Penelitian Terkait 5](#_Toc63963088)

[2.2 Part Of Speech Tagging 7](#_Toc63963089)

[2.3 Metode PoS Tagger 18](#_Toc63963090)

[2.4 Python 18](#_Toc63963091)

[2.5 *Natural Language Toolkit* 19](#_Toc63963092)

[2.6 Bahasa Indonesia 20](#_Toc63963093)

[2.7 Pengujian Teks Kalimat Metode PoS Tagger 20](#_Toc63963094)

[2.8 Pengujian Accuracy, Precision, Recall, dan F-Measure 21](#_Toc63963095)

[BAB III Metodologi Penelitian 22](#_Toc63963096)

[3.1 Bahan Penelitian 22](#_Toc63963097)

[3.2 Alat Bantu Penelitian 22](#_Toc63963098)

[3.2.1 Perangkat Keras Perangkat keras yang digunakan dalam penelitian ini adalah : 22](#_Toc63963099)

[3.2.2 Perangkat Lunak (*Software*) 22](#_Toc63963100)

[3.3 Metode Penelitian 22](#_Toc63963101)

[3.3.1 Persiapan Data 24](#_Toc63963102)

[3.3.2 Pengumpulan Tag PoS 24](#_Toc63963103)

[3.3.3 Pengembangan tag PoS 24](#_Toc63963104)

[3.3.4 Pengembangan Korpus 25](#_Toc63963105)

[3.3.5 *Tagging* Korpus 25](#_Toc63963106)

[3.3.6 Pelatihan Korpus 26](#_Toc63963107)

[3.3.7 Perancangan Pengujian 26](#_Toc63963108)

[3.3.8 Perancangan Pengujian 28](#_Toc63963109)

[3.3.9 Perancangan Pengujian 28](#_Toc63963110)

[BAB IV Hasil dan pengujian 29](#_Toc63963111)

[4.1 Hasil Penelitian 29](#_Toc63963112)

[4.1.1 Pembuatan Korpus untuk *Tagging* PoS 30](#_Toc63963113)

[4.1.3 Hasil *Tagging* PoS Metode Brigram untuk Bahasa Indonesia 30](#_Toc63963114)

[4.1.5 Hasil *Tagging* PoS Metode Trigram untuk Bahasa Indonesia 34](#_Toc63963115)

[4.1.6 Hasil *Tagging* PoS Metode Unigram untuk Bahasa Indonesia 35](#_Toc63963116)

[4.2 Pengujian 37](#_Toc63963117)

[4.2 Analisa Hasil Pengujian 38](#_Toc63963118)

[BAB V Kesimpulan dan Saran 39](#_Toc63963119)

[5.1 Kesimpulan 39](#_Toc63963120)

[5.2 Saran 39](#_Toc63963121)

[Daftar Pustaka 40](#_Toc63963122)

Daftar Tabel

[**Tabel 2. 1** Kajian Penelitian Terkait 6](#_Toc63960520)

[**Tabel 3. 1** Tabel Hasil Nilai Pengujian Accuracy, Precision, Recall, dan F-Measure 27](#_Toc63960522)

[**Tabel 4. 1** Set PoS Bahasa Melayu Pontianak 29](#_Toc60698566)

[**Tabel 4. 2** Tabel Hasil Pengujian Accuracy, Precision, Recall dan F-measure 37](#_Toc60698567)

Daftar Gambar

[**Gambar 2. 1** Set PoS Arawinda Dinakaramani (2014) 8](#_Toc60698588)

[**Gambar 2. 2** Set PoS Pisceldo(2009) Tabel 1 9](#_Toc60698589)

[**Gambar 2. 3** Set PoS Pisceldo(2009) Tabel 2 10](#_Toc60698590)

[**Gambar 2. 4** Set PoS Alfan Farizki Wicaksono (2010) 11](#_Toc60698591)

[**Gambar 2. 5** Set PoS Adriani (2009) 12](#_Toc60698592)

[**Gambar 2. 6** Set PoS Setyaningsih (2017) 13](#_Toc60698593)

[**Gambar 2. 7** *Set* PoS Purwarianti (2010) 14](#_Toc60698594)

[**Gambar 2. 8** *Set* PoS Achmad Fatchuttaman Abka (2017) 15](#_Toc60698595)

[**Gambar 2. 9** Set PoS Muliono (2012) 16](#_Toc60698596)

[**Gambar 2. 10** Set PoS Septina Dian Larasati(2011) 17](#_Toc60698597)

[**Gambar 3. 1** Diagram Alir Penelitian 23](#_Toc60698960)

[**Gambar 4. 1** Korpus latih proses tagging PoS 30](#_Toc60698964)

[**Gambar 4. 2** Proses Pelatihan IPOSTagger 30](#_Toc60698965)

[**Gambar 4. 3** Hasil Tagging Metode Bigram 32](#_Toc60698966)

[**Gambar 4. 4**Hasil Tagging Metode Perceptron 33](#_Toc60698967)

[**Gambar 4. 5**Hasil Tagging Metode Trigram 35](#_Toc60698968)

[**Gambar 4. 6** Hasil Tagging Metode Unigram 36](#_Toc60698969)

Daftar kode program

[**Kode Program 3. 1** Kode Program Pengujian Accuracy, Precision, Recall, dan F-measure 27](#_Toc60698778)

[**Kode Program 4. 1** Proses Pelatihan Bigram 31](#_Toc60698784)

[**Kode Program 4. 2** Proses Pelatihan Peceptron 32](#_Toc60698785)

[**Kode Program 4. 3** Proses Pelatihan Trigram 34](#_Toc60698786)

[**Kode Program 4. 4** Proses Pelatihan Unigram 35](#_Toc60698787)

Daftar Grafik

[**Grafik 4. 1** Hasil Pengujian Metode PoS Tagger 38](#_Toc60698809)

# BAB I Pendahuluan

## Latar Belakang

Bahasa merupakan salah satu alat komunikasi yang digunakan oleh manusia dalam kehidupan sehari-hari. Kemampuan penguasaan bahasa yang baik dapat mempermudah proses interaksi dengan orang lain. Bahasa Indonesia merupakan sistem lambang bunyi yang arbiter di gunakan oleh anggota kelompok sosial untuk berkerja sama (Krisdalaksana, 1983). Dimana setiap bangsa memilik bahasa yang berbeda-beda dengan ciri khas dan asal-usul masing-masing. Begitu juga dengan Bahasa Indonesia, Bahasa Indonesia sangat dinamis, sehingga menghasilkan kosakata yang baru dari penciptaan da penyerapan bahasa daerah maupun asing. Salah satunya bahasa yang datang dari luar adalah bahasa Inggris. Dimana Bahasa Inggris adalah bahasa Internationalyang digunakan sebagai komunikasi antar bangsa. Sehingga tidak heran, banyak orang yang belajar untuk menguasai Bahasa Inggris. Hal ini bertujuan agar mereka tidak buta akan informasi yang ada di dunia. Meskipun mempelajari Bahasa Inggris penting, akan lebih baik jika sebagai warga Negara Indonesia tetap menjaga, melestarikan dan membudayakan Bahasa kita yakni Bahasa Indonesia.

*Teknologi Text To Speech (TTS)* merupakan sebuah teknologi yang dapat membangkitkan sintesa ucapan dengan mengubah teks menjadi suara atau ucapan yang meyerupai ucapan manusia (Pelton dalam buku “Voice Processing”). TTS mengkonversi teks dalam format suatu bahasa menjadi ucapan sesuai dengan teks dalam bahasa yang digunakan. TTS sendiri dapat diimplementasikan ke dalam suatu bentuk aplikasi telekomunikasi yang dapat mempermudah pekerjaan manusia. Selain itu TTS juga dapat membantu orang-orang dengan kebutuhan khusus seperti tunanetra ataupun tunawicara.

POS (Part Of Speech) merupakan kelas kata yang dimana kelas kata ini yang dipahami oleh mesin untuk menentukan jenis dari kata-kata yang terdapat pada kalimat. PoS tag adalah suatu proses yang memberikan label kelas kata secara otomatis atau manual pada setiap kata yang ada pada suatu teks atau dokumen. Sehingga untuk pengembangan PoS Tag maka akan menggunakan Tag PoS dari beberapa penelitian sebelumnya oleh Evi Fathiyah Muniyati, Yulia Madgalena yang berjumlah 46 untuk di kembangkan dan dianalisa, sehingga mendapatkan tipe PoS yang baru dan menggunakan aplikasi iPoS Tagger dari Alfan Farizki Wicaksono (2010).

Membandingkan tipe PoS pada penelitian sebelumnya yang mana tipe PoS tersebut terdiri dari tipe PoS dari penelitian sebelumnya yaitu : tipe PoS Arawinda Dinakaramani (2014), tipe Pos Pisceldo (2009), tipe PoS Alfan Farizki Wicaksono (2010), tipe PoS Septina Dian Larasati (2011), tipe PoS Ayu Purwarianti (2010), tipe PoS Achmad Fatchuttaman Abka (2017), tipe PoS Andriani (2009), tipe PoS Setyaningsih (2017), tipe PoS Muliono (2012) yang dimana akan dibandingkan dan diuji untuk menghasilkan sebuah tipe PoS yang baru yang kemudian akan dijadikan acuan dalam penelitian kali ini.

Pada penelitian sebelumnya terdapat jumlah PoS Tagger yang memiliki jumlah lebih dari 25 yang dimana tentunya hal ini akan memberikan beban pada saat proses perhitungan dan tingkat kecepatan juga akan menurun karena harus menghitung dengan jumlah PoS Tagger yang banyak atau diatas 25. Oleh sebab itu, dalam penelitian ini, dilakukan perbandingan antara beberapa PoS Tagger kemudian dianalisa sehingga menghasilkan jumlah PoS Tagger yang berjumlah tidak lebih dari 25 dan tentunya akan dilakukan perbandingan tipe metode PoS Tagger terhadap nilai akurasi dari PoS Tagger yang telah dikembangkan.

Pengumpulan korpus yang dimana pengumpulan ini diambil dari berita online dan korpus Khamsah Akbar, pengumpulan korpus ini kemudian akan dilakukan perbaikan pada tanda baca, ejaan, atau kalimat-kalimat yang dinilai terlalu pendek atau panjang untuk pengujian, setelah ini akan dilakukan tagging PoS pada korpus menggunakan tipe PoS yang telah di analisa sebelumnya, setelah proses ini dilakukan akan dilakukan pengujian metode pos tagger terhadap pos tag yang ditelah di lakukan, kemudian akan ditarik kesimpulan, atau dilakukan pengujian kembali sesuai dengan kebutuhan yang akan dilakukan pada penelitian ini.

## Rumusan Masalah

Perumusan masalah dalam penelitian ini berdasarkan hal-hal yang telah disampaikan pada latar belakang adalah pengembangkan tipe PoS dan membandingan metode PoS Tagger untuk mencari nilai akurasi terbaik dari metode PoS Tagger.

## Tujuan Penelitian

Adapun tujuan dari penelitian yang dilakukan adalah untuk membandingan metode Pos Tagger untuk mencari nilai akurasi terbaik dari metode PoS Tagger.

0

## Pembatasan Masalah

Pembatasan masalah pada penelitian ini, sebagai berikut :

1. Korpus teks yang di gunakan adalah teks kalimat Bahasa Indonesia dari penelitian sebelumnya oleh Khamsah Akbar, dan dari beberapa berita online di Kal-Bar seperti thetanjungpuratimes.com, uncak.com, dan informatika.untan.ac.id.
2. Total Kalimat yang digunakan sebanyak 500 kalimat.
3. Bahasa Pemrograman Python dan Java,
4. Menggunakan Tools NLTK (*Natural Language Toolkit*)

## 1.5 Sistematika Penulisan

Adapun sistematika dalam penulisan tugas akhir ini tersusun atas lima bab, yang terdiri dari BAB I Pendahuluan, BAB II Tinjauan Pustaka, BAB III Metodologi Penelitian dan Perancangan Sistem, BAB IV Hasil Perancangan dan Analisis Sistem, serta BAB V Penutup.

**BAB I Pendahuluan** adalah bab pertama dalam penelitian, pada bab ini berisi latar belakang, rumusan masalah, tujuan penelitian, pembatasan masalah, sistematika penulisan.

**BAB II Tinjauan** Pustaka adalah bab yang berisi membahas tentang gambaran umum dari penelitian yang didapat dari penelitian sebelumnya serta perangkat yang melandasi pembangunan sistem dan landasan teori yang berhubungan dalam proses analisis permasalahan penelitian yang akan dilakukan.

**BAB III Metodologi Penelitian** adalah bab yang berisi tentang perangkat penelitian, metode yang akan digunakan pada penelitian, diagram alir penelitian dan perancangan pengujian yang akan dilakukan pada penelitian.

**BAB IV Hasil dan Analisis** adalah bab yang berisi penjelasan mengenai hasil dari penelitian yang dilakukan, *screenshot* tampilan antarmuka yang telah dihasilkan dan analisis pengujian yang mengarah pada suatu kesimpulan.

**BAB V Penutup** adalah bab yang berisi kesimpulan dari penelitian yang telah dilakukan dan saran atau rekomendasi untuk perbaikan, pengembangan atau kesempurnaan dan kelengkapan penelitian yang telah dilakukan.

# BAB II Tinjauan Pustaka

## 2.1 Penelitian Terkait

Beberapa penelitian terkait yang berhubungan dengan proses pemfrasaan text kalimat, Evi Fahtiyah Muniyati dengan judul skripsi *“Prediksi Jeda Dalam Ucapan Kalimat Bahasa Melayu Pontianak Menggunakan Hidden Markov Model Berbasis Part Of Speech”* dengan jeda ucapan bigram nilai akurasi frasa jeda 1+2 = 30.8% dan akurasi jeda 2 = 80% dan dengan jeda ucapan trigram nilai akurasi jeda 1+2 = 30.8% dan akurasi jeda 2 = 81.4%.

Yulia Magdalena dengan judul skripsi *“Prediksi Jeda pada Ucapan Kalimat Bahasa Melayu Pontianak Menggunakan Metode Shallow Parsing dengan Pengembangan Rule Grammar dan Rule Jeda”* dengan akurasi kecocokan pada satu kalimat sebesar 33.6% dan nilai akurasi 70,2% jeda panjang atau sebanyak 364 kalimat dari 500 kalimat.

M. Iqbal Kamiludin dengan judul skripsi *“Prediksi Jeda pada Ucapan Bahasa Melayu Pontianak dengan Menggunakan Metode Shallow Parsing”* dengan nilai recall dan precision untuk kalimat tunggal sebesar 0.78 dan 0.74, sedangkan untuk kalimat majemuk sebesar 0.67 dan 0.57. Dari 168 kalimat yang ada, nilai kecocokan dengan jeda penutur sebesar 40.4% atau 68 kalimat.

Adhitya Teguh Nugraha (2014), Teknik Informatika di Universitas Tanjungpura dalam skripsi yang berjudul Prediksi Jeda dalam Ucapan Kalimat Bahasa Indonesia dengan *Hidden Markov Model* menjelaskan bahwa tujuan penelitian tersebut adalah memprediksi terjadinya jeda pada ucapan kalimat berbahasa Indonesia dengan menggunakan *Hidden Markov Model* (HMM). Penelitian ini menghasilkan aplikasi untuk memprediksi jeda dalam ucapan kalimat berbahasa Indonesia dan menggunakan Delphi untuk membuat antarmuka sistem. Korpus yang digunakan berasal dari berita yang disiarkan oleh Lembaga Penyiaran Publik TVRI Kalimantan Barat. Penelitian ini menggunakan 35 *tag* PoS seperti yang digunakan pada penelitian milik Wicaksono dan Purwarianti (2010). Hasil dari penelitian ini yaitu nilai *recall* sebesar 0.132, nilai *precision* sebesar 0.364 dan *f-score* sebesar 0.194.

Dari penelitian sebelumnya dapat dilihat dari nilai akurasi dan kecocokan nilai akurasi yang paling tinggi didapatkan dari penelitian oleh Evi Fahtiyah Muniyati dengan Tag PoS yang berjumlah 45, dalam hal ini dapat di kembangkan lagi dengan mengurangi jumlah PoS tag pada penelitian sebelumnya.

**Tabel 2. 1** Kajian Penelitian Terkait

| No. | Penulis | Metode | Hasil |
| --- | --- | --- | --- |
| 1 | Evi Fahtiyah Muniyati, Prediksi Jeda Dalam Ucapan Kalimat Bahasa Melayu Pontianak Menggunakan Hidden Markov Model Berbasis Part Of Speech, 2019. | * Menggunakan *tools* HMM yaitu IPOSTagger. | * jeda ucapan bigram nilai akurasi frasa jeda 1+2 = 30.8% dan akurasi jeda 2 = 80% * jeda ucapan trigram nilai akurasi jeda 1+2 = 30.8% dan akurasi jeda 2 = 81.4% |
| 2 | Yulia Magdalena, Prediksi Jeda pada Ucapan Kalimat Bahasa Melayu Pontianak Menggunakan Metode Shallow Parsing dengan Pengembangan Rule Grammar dan Rule Jeda, 2019. | * Metode yang digunakan *Shallow Parsing* * Menggunakan pengembangan *Rule Grammar*  dan *Rule Jeda* | * akurasi kecocokan pada satu kalimat sebesar 33.6% dan nilai akurasi 70,2% jeda panjang atau sebanyak 364 kalimat dari 500 kalimat. |
| 3 | M. Iqbal Kamiludin, Prediksi Jeda pada Ucapan Bahasa Melayu Pontianak dengan Menggunakan Metode *Shallow Parsing*, 2017. | * Metode yang digunakan *Shallow Parsing*. * Menggunakan *rule grammar* untuk memprediksi jeda. | * Nilai *recall* sebesar 0.78, *precision* sebesar 0.74 dan *f-measure* sebesar 0.75 untuk kalimat tunggal. * Nilai *recall* sebesar 0.67, *precision* sebesar 0.57 dan *f-measure* sebesar 0.61 untuk kalimat majemuk. |
| 4 | Adhitya Teguh Nugraha, Prediksi Jeda dalam Ucapan Kalimat Bahasa Indonesia dengan *Hidden Markov Model*, 2014. | * Menggunakan *tools* HMM yaitu IPOSTagger. * GUI aplikasi menggunakan Delphi. | * Nilai *recall* sebesar 0.132, *precision* sebesar 0.364 dan *f-score* sebesar 0.194. |

Penelitian ini dilakukan untuk mengetahui nilai akurasi PoS Tag yang telah dikembangkan terhadap metode PoS Tagger pada Bahasa Indonesia.

## 2.2 Part Of Speech Tagging

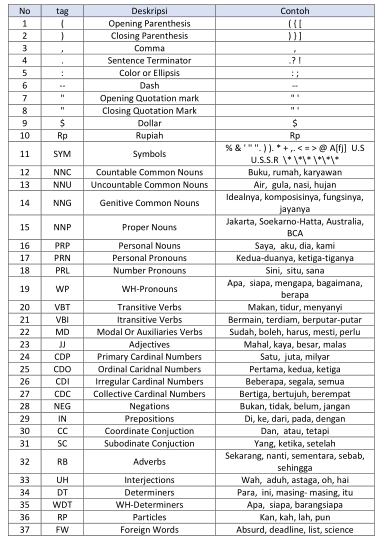
*Part of speech* (PoS) *tagging* adalah proses menentukan kelas kata untuk setiap kata dalam sebuah kalimat. *Part of speech* atau kelas kata terdiri dari kata kerja, kata sifat, kata benda, kata keterangan, dan lain-lain. Beberapa penggunaan PoS-*tagging* adalah untuk menghapus perbedaan yang tidak relevan, menghapus ambiguitas, membantu *stemming,* dan membantu pencarian kata benda (Christianti M, Pragantha dan Victor, 2016).

Hingga saat ini sudah ada beberapa *set* PoS berbahasa Indonesia yang dihasilkan dari penelitian sebelumnya, diantaranya yaitu tipe PoS Arawinda Dinakaramani(2014), tipe Pos Pisceldo(2009), tipe PoS Alfan Farizki Wicaksono(2010), tipe PoS Septina Dian Larasati(2011), tipe PoS Ayu Purwarianti(2010), tipe PoS Achmad Fatchuttaman Abka(2017), tipe PoS Andriani(2009), tipe PoS Setyaningsih(2017), tipe PoS Muliono(2012). Kesepuluh penelitian tersebut dijadikan sebagai acuan untuk pengembangan *set* PoS pada penelitian ini.

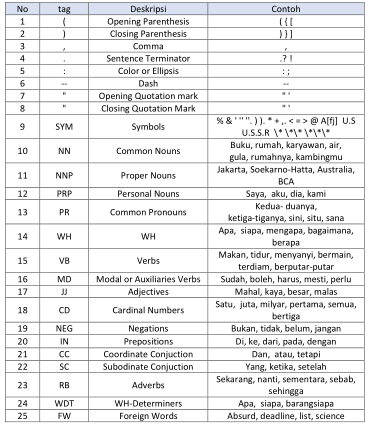
Kesepuluh penelitian tersebut dijadikan sebagai acuan untuk pengembangan *set* PoS pada penelitian ini.



**Gambar 2. 1** Set PoS Arawinda Dinakaramani (2014)



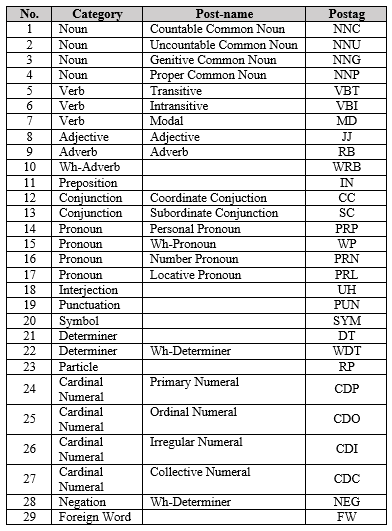
**Gambar 2. 2** Set PoS Pisceldo(2009) Tabel 1



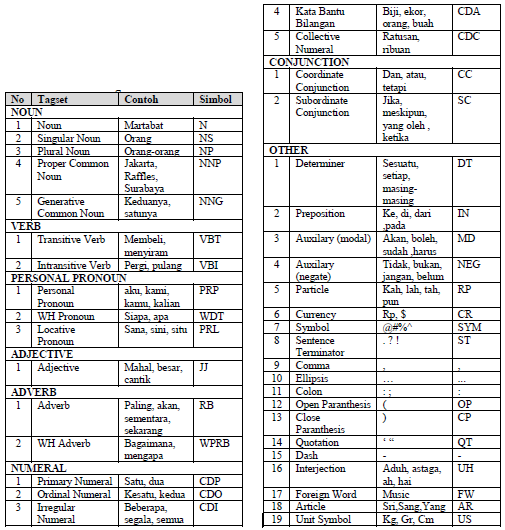
**Gambar 2. 3** Set PoS Pisceldo(2009) Tabel 2



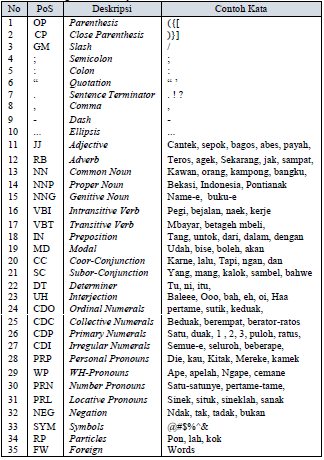
**Gambar 2. 4** Set PoS Alfan Farizki Wicaksono (2010)



**Gambar 2. 5** Set PoS Adriani (2009)



**Gambar 2. 6** Set PoS Setyaningsih (2017)



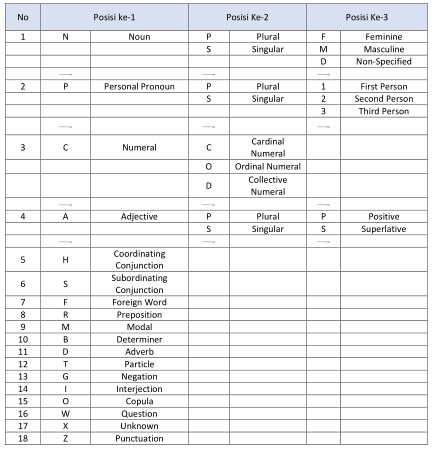
Gambar 2. 7 *Set* PoS Purwarianti (2010)



Gambar 2. 8 *Set* PoS Achmad Fatchuttaman Abka (2017)



**Gambar 2. 9** Set PoS Muliono (2012)



**Gambar 2. 10** Set PoS Septina Dian Larasati(2011)

Berdasarkan gambar tagging PoS tersebut maka penelitian ini akan membuat Tagging PoS Bahasa Indonesia yang baru berdasarkan referensi Tag PoS yang sudah dijelaskan untuk keperluan perhitungan nilai akurasi metode PoS Tagger.

## 2.3 Metode PoS Tagger

Metode PoS Tagger yang digunakan pada penelitian ini adalah, Bigram, Unigram, Trigram, HMM, Perceptron. Penjelasan mengenai metode tagger adalah sebagai berikut :

Unigram

Unigram tagger adalah sebuah tagger yang hanya menggunakan satu kata sebagai konteksnya untuk menentukan Tag PoS(*Part Of Speech*), sederhananya tagger ini berbasis konteks yang konteksnya adalah satu kata, yaitu Unigram, yang dimana n = 1.

Bigram

Brigram tagger adalah pada dasarnya bekerja dengan cara yang sama seperti unigram, perbedaannya adalah pada bigram tagger mempertimbangkan konteks kapan menetapkan tag ke kata, sehingga jika dijelaskan lebih rinci bigram adalah urutan dua eleman yang berdekatan dari untaian token yang biasanya berupa huruf, suku kata, atau kata yang dimana n = 2.

Trigram

Trigram tagger adalah sebuah tagger yang menggunakan dua kata sebagai konteksnya untuk menentukan Tag PoS(*Part Of Speech*), yang dimana n = 3.

HMM (Hidden Markov Model)

HMM (*Hidden Markov Model*) dikembangkan dari *Markov Chain* yaitu keadaan yang akan datang dari suatu *sequence* tidak ditentukan dari keadaan saat ini, tapi juga perpindahan dari suatu *state* ke *state* *sequence* yang lain.

Perceptron

Perceptron tagger mengimplementasikan penandaan PoS (*Part Of Speech*) menggunakan rata-rata, struktur algoritma perceptron.

## 2.4 Python

Menurut Ljubomir Perkovic (2012), Python merupakan bahasa pemrograman dengan tujuan umum yang dikembangkan secara khusus untuk membuat source code mudah dibaca. Python juga memiliki library yang lengkap sehingga memungkinkan *programmer* untuk membuat aplikasi yang mutakhir dengan menggunakan *source code* yang tampak sederhana.

Beberapa keunggulan Python apabila dibandingkan dengan bahasa pemrograman lain adalah :

* Memiliki kepustakaan yang luas, dalam distribusi Python telah disediakan modul-modul `siap pakai’ untuk berbagai keperlua.
* Memiliki tata bahasa yang jernih dan mudah dipelajari.
* Memiliki aturan layout kode sumber yang memudahkan pengecekan, pembacaan kembalidan penulisan ulang kode sumber.
* Berorientasi obyek.
* Memiliki sistem pengelolaan memori otomatis (*garbage collection*, seperti java) Modullar, mudah dikembangkan dengan menciptakan modul-modul baru; modul-modul tersebut dapat dibangun dengan bahasa Python maupun C/C++.

Salah satu kegunaan Bahasa Pemrograman Python yaitu dapat mendukung aktivitas pemrosesan Basahasa alami, dengan menggukan modul dari *Natural Language Toolkit* yang merupakan salah satu koponen yang disediakan oleh Bahasa pemrograman python, khususnya dalam pengolahan digital Bahasa alami. Dalam penelitian ini python digunakan sebagai Bahasa pemrograman dalam proses menentukaan frasa jeda ucapan pada kalimat dengan menggunakan fungsi *shallow* parsing yang merupakan salah satu tugas dari Natural Language Toolkit.

## 2.5 Natural Language Toolkit

Walaupun python memiliki kemampuan untuk melakukan tugas-tugas NLP (*Natural Language Processing*) secara dasar, namun tidak cukup efisien dalam melakukan tugas-tugas standar yang terdapat dalam NLP, maka dari itu digunakan modul NLTK (*Natural Language Toolkit)*. Modul NLTK menyediakan berbagai fungsi dan *wrapper,* serta *corpora* standar baik itu mentah atau *pre-processed* yang digunakan dalam materi pengajaran NLP. NLTK sendiri merupakan sebuah *platform* yang dirilis oleh Steven Bird dan Edward Loper pada tahun 2001 yang digunakan untuk membangun program analisis teks. NLTK menyediakan antarmuka yang mudah digunakan untuk lebih dari 50 sumber korpora dan leksikal dan seperangkat pustaka (*library*) pemrosesan teks untuk klasifikasi (*classification*), tokenisasi (*tokenization*), penumpukan (*stemming*), penandaan (*tagging*), penguraian (*parsing*), penalaran semantik (*semantic reasoning*), pembungkus (*wrapper*) pustaka NLP dan forum diskusi aktif (Lestari, 2017).

## 2.6 Bahasa Indonesia

Bahasa Indonesia adalah sistem lambang bunyi yang arbiter digunakan oleh anggota kelompok sosial untuk bekerja sama (Kridalaksana, 1983). Dimana setiap bangsa memiliki bahasa yang berbeda-beda dengan ciri khas dan asal-usul masing-masing. Begitu juga dengan bahasa Indonesia. Sejarah bahasa Indonesia sendiri tidak lepas dari Bahasa Indonesia. Bahasa Indonesia sangat dinamis, sehingga menghasilkan kosakata baru dari penciptaan dan penyerapan bahasa daerah maupun asing. Salah satu bahasa yang datang dari luar adalah bahasa Inggris. Dimana bahasa Inggris adalah bahasa internasional yang digunakan sebagai komunikasi antar bangsa. Sehingga tidak heran, banyak orang yang belajar untuk menguasai bahasa Inggris. Hal ini bertujuan agar mereka tidak buta akan informasi yang ada di dunia. Meskipun mempelajari bahasa Inggris penting, akan lebih baik jika sebagaiwarga Negara Indonesia tetap menjaga, melestarikan dan membudayakan bahasa Indonesia.

## 2.7 Pengujian Teks Kalimat Metode PoS Tagger

Pengujian teks kalimat yang lakukan antara kalimat tagging manual yaitu korpus yang telah dikembangkan kemudian dilakukan tagging terlebih dahulu dengan cara manual yaitu sebagai contoh :

“Pada hari Sabtu, delapan februari tahun dua ribu dua puluh telah dilaksanakan Uji Sertifikasi Kompetensi untuk dosen-dosen Jurusan Informatika UNTAN .”

Kalimat diatas adalah contoh kalimat dari korpus yang telah dikembangakn sehingga berjumlah 500 kalimat, kemudian dilakukan tagging dengan menambahkan “/” yaitu garis miring dan label kelas kata yaitu tag PoS, sehingga menjadi “Pada/NN hari/NN Sabtu/NNP ,/, delapan/CD februari/NN tahun/NN dua/CD ribu/CDA dua/CD puluh/CDA telah/VB dilaksanakan/VB Uji/NN Sertifikasi/NNP Kompetensi/NNP untuk/IN dosen-dosen/RB Jurusan/NNP Informatika/NNP UNTAN/NNP ./.” hal ini perlu dilakukan agar mesin mengetahui apa jenis kata yang digunakan atau kelas kata yang terdapat pada kalimat tersebut, hal ini dilakukan sebanyak 500 kalimat semuanya dilakukan tagging secara manual, kemudian akan dilakukan perbandingan antara tagging manual tersebut dengan tagging metode PoS Tagger.

## 2.8 Pengujian Accuracy, Precision, Recall, dan F-Measure

*Precision* dan *recall* menjadi ukuran yang paling sering digunakan untuk mengevaluasi suatu sistem. *Precision* adalah rasio jumlah dokumen relevan yang ditemukan dengan total jumlah dokumen yang ditemukan oleh sistem (Rozi, Pramono dan Dahlan, 2012). Nilai *precision* tertinggi adalah 1, yang berarti seluruh dokumen yang ditemukan adalah relevan (Fitri, 2013). *Recall* adalah rasio jumlah dokumen relevan yang ditemukembalikan terhadap jumlah seluruh dokumen relevan dalam koleksi (Nugraha, 2014). Nilai *recall* tertinggi adalah 1, yang berarti seluruh dokumen dalam koleksi berhasil ditemukan. Imam Fahrur Rozi (2012) menjelaskan bahwa, precision adalah rasio jumlah dokumen relevan yang ditemukan dengan total jumlah dokumen yang ditemukan oleh sistem. Recall adalah rasio jumlah dokumen relevan yang ditemukan kembali dengan total jumlah dokumen dalam kumpulan dokumen yang dianggap relevan. Precision dapat dianggap sebagai ukuran ketepatan atau ketelitian, sedangkan recall adalah kesempurnaan. Accuracy merupakan nilai dari keakuratan hasil pengujian dari yang telah di jalankan. Accuracy di defefnisikan sebagai tingkat kedekatan antara nilai prediksi dengan nilai aktual.

F-measure (nilai F atau F1-Score atau score) 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.

Pengujian *accuracy*, *precision*, *recall*, dan *f-measure* digunakan untuk mengetahui nilai akurasi keberhasilan PoS Tag terhadap beberapa metode PoS Tagger.

# BAB III Metodologi Penelitian

* 1. **Bahan Penelitian**

Bahan penelitian berupa Korpus dari berita online thetanjungpuratimes.com, uncak.com, informatika.intan.ac.id, dan korpus Bahasa Indonesia sebelumnya dari Khamsah Akbar.

* 1. **Alat Bantu Penelitian**

Perangkat penelitian yang digunakan dalam penelitian ini berupa perangkat lunak dan perangkat keras. Adapun penjabaran perangkat penelitian yang digunakan adalah sebagai berikut :

* + 1. **Perangkat Keras**  
       Perangkat keras yang digunakan dalam penelitian ini adalah :

1. Laptop ASUS X505Z, AMD Dual Core R3-2200U, Kapasitas RAM 4GB, HDD 1 TB
   * 1. **Perangkat Lunak (*Software*)**

Perangkat lunak yang digunakan dalam penelitian ini adalah :

1. Sistem Operasi Windows 10
2. Python 3.8.1 dan Phycharm Community
3. NLTK tools
4. IPOSTagger sebagai *tools* yang digunakan untuk *tagging* PoS
5. Sublime Text 3 sebagai aplikasi untuk teks *editor.*
   1. **Metode Penelitian**

Metode Penelitian yang akan dilakukan dijelaskan pada diagram alir penelitian pada penelitian 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 pengumpulan tag pos, membandingkan tag pos, pengembangan pos, pembuatan korpus tagging pos, pelatihan korpus, ujicoba korpus, pengujian, analisa hasil pengujian, dan kesimpulan.

* + 1. **Persiapan Data**

Persiapan data berupa teks kalimat bahasa Indonesia yang bersumber dari berita online thetanjungpuratimes.com, uncak.com, informatika.untan.ac.id, dan korpus Bahasa Indonesia sebelumnya dari Khamsah Akbar. Jumlah kalimat yang digunakan pada penelitian sebanyak 500 kalimat. Teks kalimat digunakan untuk penandaan PoS atau label kata.

* + 1. **Pengumpulan Tag PoS**

Pengumpulan tag PoS dari penelitian sebelumnya, yang dimana penelitian proses pengumpulan tipe PoS ini terdari dari 10 penelitian sebelumnya, yaitu : tipe PoS Arawinda Dinakaramani(2014), tipe Pos Pisceldo(2009), tipe PoS Alfan Farizki Wicaksono(2010), tipe PoS Septina Dian Larasati(2011), tipe PoS Ayu Purwarianti(2010), tipe PoS Achmad Fatchuttaman Abka(2017), tipe PoS Andriani(2009), tipe PoS Setyaningsih(2017), tipe PoS Muliono(2012).

* + 1. **Pengembangan tag PoS**

Pada tahap ini digunakan set PoS yang baru dibentuk berdasarkan Set PoS Bahasa Indonesia. Pengelompokkan kelas kata dalam Bahasa Indonesia dilakukan secara manual dengan merujuk pada referensi tipe PoS Arawinda Dinakaramani(2014), tipe Pos Pisceldo(2009), tipe PoS Alfan Farizki Wicaksono(2010), tipe PoS Septina Dian Larasati(2011), tipe PoS Ayu Purwarianti(2010), tipe PoS Achmad Fatchuttaman Abka(2017), tipe PoS Andriani(2009), tipe PoS Setyaningsih(2017), tipe PoS Muliono(2012). Set PoS baru dibuat dengan tujuan untuk memperjelas penandaan tipe kata dalam kalimat dan meningkatkan nilai akurasi. Set PoS ini digunakan pada proses pengembangan rule grammar yang digunakan pada proses shallow parsing pada tahap selanjutnya. Pembuatan tabel PoS bahasa Indonesia dilakukan dengan cara menganalisa tabel PoS referensi yang ada dan mencocokkan label PoS yang sesuai dengan kelas kata yang digunakan pada bahasa Indonesia. Penambahan set PoS disesuaikan dengan fungsi kata dalam kalimat bahasa Indonesia. Misalkan, pada penelitian Pada penelitian Setyaningsih (2017) , terdapat label PoS AR atau artikel untuk panggilan seperi “si”, “sang”, atau “Yang”, sehingga PoS AR diambil dan ditambahkan ke tabel PoS yang baru. Hasil dari pembuatan tabel PoS adalah tabel set PoS baru yang digunakan untuk menandai kata di korpus teks bahasa Indonesia.

* + 1. **Pengembangan Korpus**

Korpus dari berita online thetanjungpuratimes.com, uncak.com, informatika.intan.ac.id, dan korpus Bahasa Indonesia sebelumnya dari Khamsah Akbar, sehingga dari pengembangan ini mendapatkan total kalimat sebesar 500 kalimat teks bahasa Indonesia, sehingga korpus menjadi lebih maksimal untuk pengujian.

* + 1. ***Tagging* Korpus**

Pada dasarnya, IPOSTagger menggunakan korpus Bahasa Indonesia. Teks kalimat untuk korpus latih IPOSTagger yang berasal dari berita online thetanjungpuratimes.com, informatika.untan.ac.id, uncak.com, dan corpus sebelumnya dari Khamsah Akbar 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”. Korpus latih tersebut berisikan “kata/PoS”.

Misalkan pada korpus terdapat kalimat :

“eh, kamu jangan banyak bicara .”

Setiap kata akan dilihat dan dikelompokkan berdasarkan label PoS. Misalkan kata “Eh” menunjukkan panggilan pada kalimat dengan kata seru sehingga ditandai dengan PoS “RP” untuk kata seru. Kata “kamu” merupakan objek yang berupa kata ganti orang sehingga di tandai dengan PoS “PRP”, “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 “bicara” menunjukkan kegiatan berbicara sehingga ditandai dengan PoS “VB” untuk kata kerja intrasitif yang memerlukan objek dibelakang kata kerja. Hasil penandaan PoS menjadi kalimat yang berisi “kata/PoS” seperti berikut:

“Eh/RP ,/, kamu/PRP jangan/NEG banyak/JJ bicara/VB ./.“

* + 1. **Pelatihan Korpus**

Pelatihan korpus dilakukan menjadi dua bagian yaitu dengan menggunakan aplikasi IPOSTagger untuk metode HMM, dan menggunakan aplikasi PyCharm Community untuk pelatihan metode Unigram, Bigram, Trigram, dan Perceptron.

##### 3.3.6.1 Pelatihan Korpus di IPOSTagger

Setelah korpus latih selesai dibuat, selanjutnya dilakukan proses pelatihan korpus pada IPOSTagger untuk keperluan implementasi prediksi jeda. Pada proses pelatihan korpus untuk tagging PoS, korpus latih yang berisi “kata/PoS” disimpan dalam folder IPOSTagger dengan ekstensi \*.crp. Kemudian proses pelatihan dilakukan melalui command prompt pada folder IPOSTagger.

##### 3.3.6.2 Pelatihan Korpus Unigram, Bigram, Trigram, Perceptron.

Setelah dilakukan pelatiahn korpus pada IPOSTagger yaitu pelatihan metode HMM, maka dilakukan pelatihan pada metode *PoS Tagger* lainnya. Pada pelatihan korpus untuk tagging PoS, korpus latih yang berisi “kata/PoS” disimpan pada folder masing-masing metode dengan nama file uji500.txt. kemudian proses pelatihan dilakukan melalui aplikasi phycharm community.

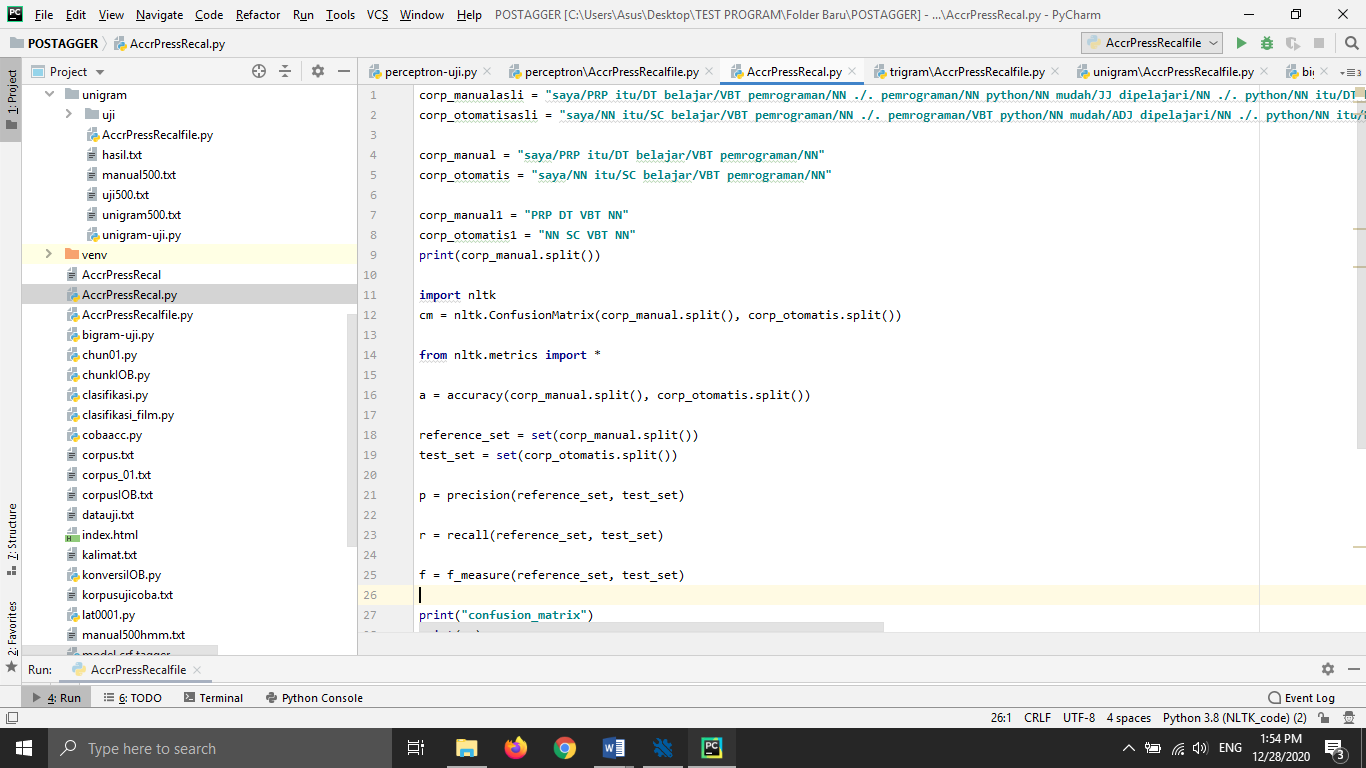
**3.3.7 Perancangan Pengujian**

Pengujian dilakukan untuk mendapatkan nilai pengujian yang akan dianalisis menjadi kesimpulan penelitian. Pengujian dilakukan dengan membandingkan kalimat asli dengan kalimat latih dari hasil implementasi menggunakan pengujian Precision, Recall dan F-Score dan pengujian akurasi kecocokan dalam satu kalimat. Total kalimat yang akan diuji adalah 500 kalimat bahasa Indonesia.

##### 3.3.7.1 Skenario Pengujian *Accuracy*, *Recall*, *Precision*, dan *F-Measure*

Pada pengujian ini, dilakukan perhitungan dari setiap korpus uji dari setiap metode yang telah dilatih, dan kemudian pengujian dijalan degan menjalan aplikasi phycharm community dengan koding yang telah dibuat dan dijalankan untuk mendapatkan hasil dari nilai setiap pengujian. Kode program untuk setiap pengujian adalah sebagai berikut :

**Kode Program 3. 1** Kode Program Pengujian Accuracy, Precision, Recall, dan F-measure



Setelah dilakukan pengujian menggunakan koding pada gambar 3.2 hasil yang di hasilkan akan di tulis pada tabel 3.1 yang dimana hasil ini lah yang menjadi hasil dari pengujian metode PoS Tagger ini.

**Tabel 3. 1** Tabel Hasil Nilai Pengujian Accuracy, Precision, Recall, dan F-Measure

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Metode PoS Tagger | Accuracy\_Score | Precision\_Score | Recall\_Score | F-measure |
| Bigram | Nilai akurasi | Nilai akurasi | Nilai akurasi | Nilai akurasi |
| Metode PoS Tagger | Accuracy\_Score | Precision\_Score | Recall\_Score | F-measure |
| HMM | Nilai akurasi | Nilai akurasi | Nilai akurasi | Nilai akurasi |
| Perceptron | Nilai akurasi | Nilai akurasi | Nilai akurasi | Nilai akurasi |
| Trigram | Nilai akurasi | Nilai akurasi | Nilai akurasi | Nilai akurasi |
| Unigram | Nilai akurasi | Nilai akurasi | Nilai akurasi | Nilai akurasi |

**3.3.8 Perancangan Pengujian**

Pada tahap ini, hasil pengujian berupa nilai akurasi dalam satu kalimat

penuh dan nilai precision, recall, dan f-measure dianalisis secara keseluruhan untuk

mempermudah penarikan kesimpulan.

**3.3.9 Perancangan Pengujian**

Kesimpulan dirumuskan berdasarkan dari perbandingan nilai akurasi dari setiap metode PoS Tagger, yang dimana setiap nilai dari metode ini berpengaruh dari seberapa tinggi nilai akurasi antara setiap metode yang diuji.

# BAB IV Hasil dan pengujian

## Hasil Penelitian

Hasil dari penelitian yang dilakukan yaitu berupa nilai akurasi dari metode PoS Tagger untuk Bahasa Indonesia.

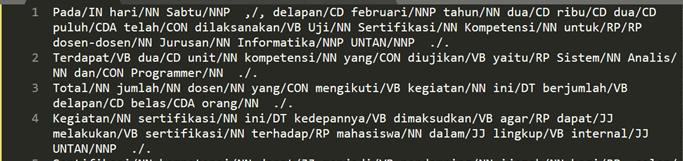
Hasil pengembangan Tag PoS bahasa Indonesia yaitu tabel PoS dengan jumlah sebanyak 22 set PoS. Set PoS bahasa Indonesia dapat dilihat pada tabel 4.1.

**Tabel 4. 1** Set PoS Bahasa Melayu Pontianak

|  |  |  |  |
| --- | --- | --- | --- |
| No | tag | Deskripsi | Contoh |
| 1 | NN | nouns | kata benda |
| 2 | NNP | proper nouns | nama, nnama benda atau nama orang, |
| 3 | PRP | Personal Nouns | Saya, aku, dia, kami, orang |
| 4 | VB | Verbs | Makan, tidur, menyanyi, bermain, terdiam, berputar-putar |
| 5 | JJ | adjective | kata sifat |
| 6 | RB | adverb | kata keterangan |
| 7 | CD | Cardinal Number | ratusan, ribuan, number, satu, dua, kesatu, ketiga, angka bilangan Rp, $, Kedua- duanya, ketiga-tiganya, sini, situ, sana |
| 8 | CDA | Kata Bantu Bilangan | Biji, Ekor, Orang, Buah orang, ton, helai, lembar |
| 9 | CDI | Irregular Numeral | Beberapa, Segala, Semua, BANYAK |
| 10 | CON | Conjunction | Dan, Atau, Tetapi, sejak, jika, seandainya, supaya, meski, seolah-olah, sebab, maka, tanpa, dengan, bahwa, yang, lebih ... daripada ..., semoga |
| 11 | DT | Determiners | Para, ini, masing- masing, itu, sang Sri, Sang, Yang, ini, itu, sini, situ |
| 12 | IN | Preposition | ke, di, dari, pada, tentang |
| 13 | MD | Auxiliary(modal) | akan, boleh, sudah, harus, mesti, Bisa |
| 14 | RP | Particle | Kah, lah, tah, pun, Aduh, astaga, ah, hai, oi |
| 15 | WH | WH-Determiners | Apa, siapa, barangsiapa,Apa, siapa, mengapa, bagaimana, berapa |
| 16 | NEG | Negations | Bukan, tidak, belum, jangan |
| 17 | FW | Foreign Word | Music, tidak ketahui |
| 18 | SYM | Symbol | @#%^ , …, : ; , ( ) {}[]\ <> - + = ` ~ Kg, Gr, Cm , ..., ?, . |
| 19 | . | TITIK | . |
| 20 | , | KOMA | , |
| 21 | ? | Tanda Tanya | ? |
| 22 | ! | Tanda seru | ! |

### Pembuatan Korpus untuk *Tagging* PoS

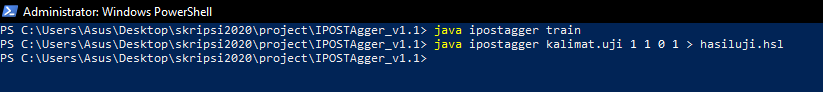
Korpus dasar IPOSTagger yang berbahasa Indonesia. Sebanyak 500 teks kalimat yang digunakan sebagai korpus latih yang berasal dari berita online ditandai kelas katanya secara manual. Korpus latih untuk *tagging* PoS yang berisi “kata/PoS” dapat dilihat pada gambar 4.1.



**Gambar 4. 1** Korpus latih proses tagging PoS

* + 1. **Hasil *Tagging* PoS Metode HMM untuk Bahasa Indonesia**

Selanjutnya, teks kalimat bahasa Indonesia menggunakan set PoS baru yang sudah dikembangkan. Penandaan atau *tagging* PoS bertujuan untuk mencari nilai akurasi dari metode hmm. Korpus yang telah diberi label PoS dapat dilihat pada gambar 4.1



**Gambar 4. 2** Proses Pelatihan IPOSTagger

Tiap kata dalam kalimat teks bahasa Indonesia diberi tanda”/” atau garis miring dan diikuti label PoS, penandaan dilakukan secara manual yang telah dilakukan seperti pada gambar 4.1 kemudian di simpan pada subfolder “*resource*” pada folder IPOSTagger dengan extensi .crp. Kemudian proses pelatihan dilakukan melalui *command prompt* pada folder IPOSTagger seperti pada gambar 4.2.

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

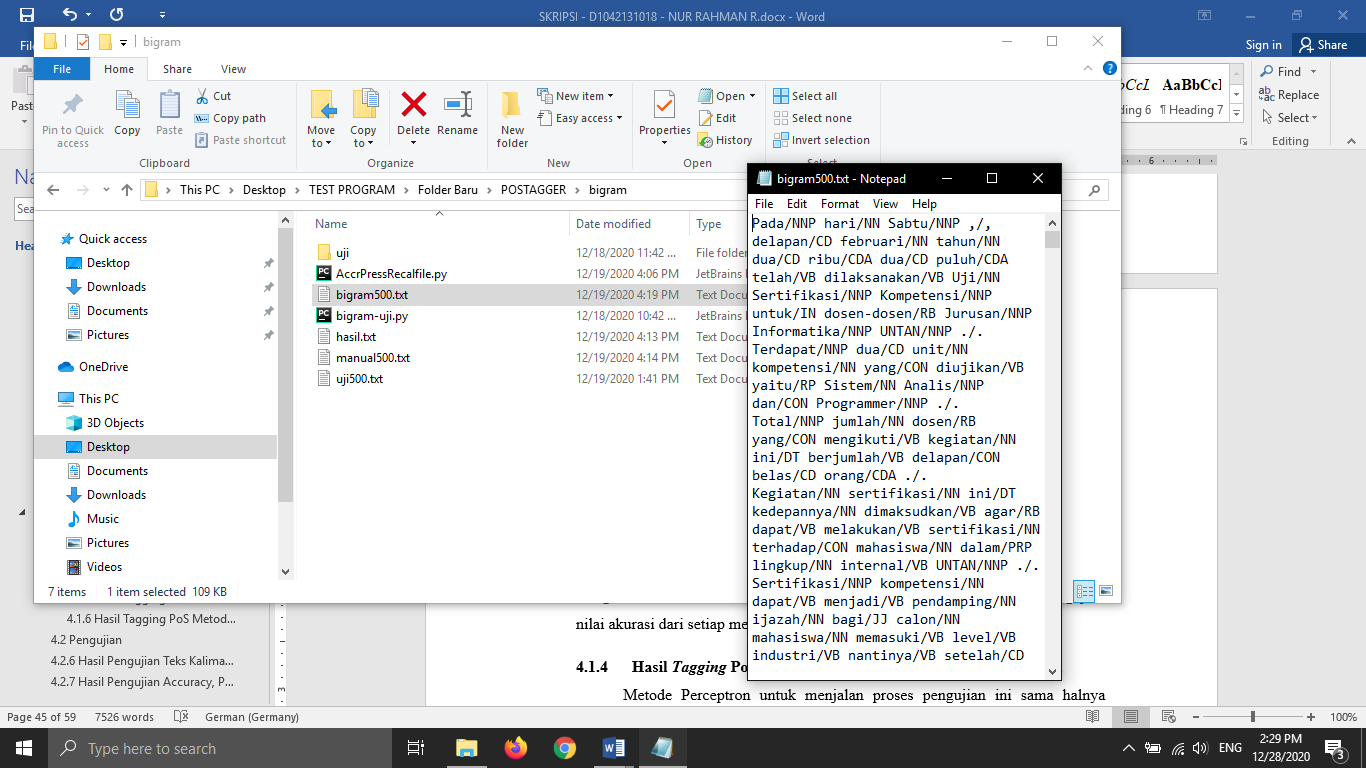
### Hasil *Tagging* PoS Metode Brigram untuk Bahasa Indonesia

Teks kalimat bahasa Indonesia kemudian dilakukan pengujian terhadap metode lainnya yaitu metode *bigram* yang dimana kalimat yang digunakan untuk latih dan pengujian sama seperti pada pengujian metode *hmm*, namun yang berbeda pada pengujian ini adalah proses melakukan pengujiannya yang dimana sebelumnya menggunakan command prompt, pada kali ini menggunakan aplikasi *phycharm community*, koding untuk menjalan proses ini, adalah sebagai berikut :

**Kode Program 4. 1** Proses Pelatihan Bigram

import nltk  
import bigram  
import time  
startTime = time.time()  
  
*#kalimat uji*kalimatuji=[]  
f1 = open("uji500.txt", "r")  
for baris in f1 :  
 kaluji = baris.strip()  
 kalimatuji.append(kaluji.split())  
 *#print(kalimatuji, len(kalimatuji))*train\_data=[]  
f0 = open("manual500.txt", "r") *#corpustrain.txt*for baris in f0 :  
 kal = baris.strip()  
 kalimat =[nltk.tag.str2tuple(t) for t in kal.split()]  
 *#print(kalimat)* train\_data.append(kalimat)  
  
*#print(train\_data)*databigram=[]  
bigram\_tagger = nltk.BigramTagger(train\_data)  
for jumkal in range(len(kalimatuji)):  
 *#print("Unigram =", unigram\_tagger.tag(kalimatuji[jumkal]))* databigram.append([nltk.tag.tuple2str(x) for x in bigram\_tagger.tag(kalimatuji[jumkal])])  
  
  
x=0  
for x in range(0, len(databigram)):  
 y=0  
 txtOut = ""  
 for y in range(0, len(databigram[x])):  
 txtOut=txtOut + databigram[x][y]+ ' '  
 print(txtOut [:-1])  
"""oleh clik jawaban dulu lalu proses durasi --> stop time"""  
print(f"waktu eksekusi = {time.time()-startTime}")

Setelah menjalan koding ini pada aplikasi *phycharm community* maka kita akan mendapat hasil yang seperti seperti pada gambar 4.1 namun isinya kan berbeda karena setiap metode memberikan hasil yang berbeda dan hasilnya adalah sebagai berikut :



**Gambar 4. 3** Hasil Tagging Metode Bigram

Seperti yang dapat dilihat terdapat beberapa perbedaan antara gambar 4.1 dan gambar 4.3 ini, data ini lah yang akan digunakan sebagai bahan untuk pengujian nilai akurasi dari setiap metode PoS Tagger.

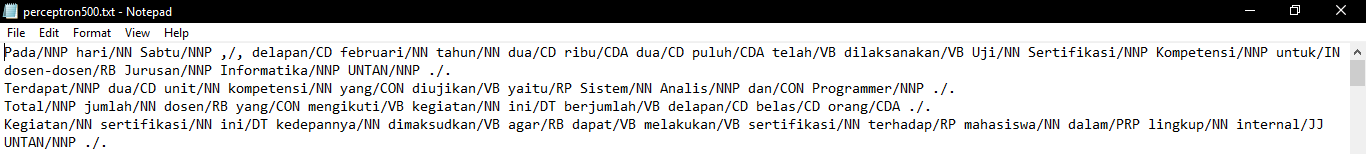
**4.1.4 Hasil *Tagging* PoS Metode Perceptron untuk Bahasa Indonesia**

Metode *Perceptron* untuk menjalan proses pengujian ini sama halnya dengan metode *bigram*, tetapi tetap hasil keluaran dari metode ini tentunya akan berbeda juga dengan *hmm*, dan *bigram*. Dibawah ini adalah koding untuk proses *tagging* metode *perceptron*.

**Kode Program 4. 2** Proses Pelatihan Peceptron

import perceptron  
from nltk.tag.perceptron import PerceptronTagger  
import time  
startTime = time.time()  
  
*#kalimat uji*kalimatuji=[]  
f1 = open("uji500.txt", "r")  
for baris in f1 :  
 kaluji = baris.strip()  
 kalimatuji.append(kaluji.split())  
 *#print(kalimatuji, len(kalimatuji))*train\_data=[]  
f0 = open("manual500.txt", "r") *#corpustrain.txt*for baris in f0 :  
 kal = baris.strip()  
 kalimat =[nltk.tag.str2tuple(t) for t in kal.split()]  
 *#print(kalimat)* train\_data.append(kalimat)  
  
  
*#print(train\_data)*dataperception=[]  
tagger = PerceptronTagger(load=True)  
tagger.train(train\_data)  
for jumkal in range(len(kalimatuji)):  
 *#print("perception =", tagger.tag(kalimatuji[jumkal]))  
 # dataperception.append(tagger.tag(kalimatuji[jumkal]))* dataperception.append([nltk.tag.tuple2str(x) for x in tagger.tag(kalimatuji[jumkal])])  
  
*#cetak data*x=0  
for x in range(0, len(dataperception)):  
 y=0  
 txtOut = ""  
 for y in range(0, len(dataperception[x])):  
 txtOut=txtOut + dataperception[x][y]+ ' '  
 print(txtOut [:-1])  
"""oleh clik jawaban dulu lalu proses durasi --> stop time"""  
print(f"waktu eksekusi = {time.time()-startTime}")

Setelah dilakukan koding seperti diatas maka hasilnya akan seperti pada gambar 4.4 yang dimana hasil *tagging* dari metode *perceptron* ini tentunya akan berbeda dengan hasil dari metode *hmm*, *bigram*.



**Gambar 4. 4**Hasil Tagging Metode Perceptron

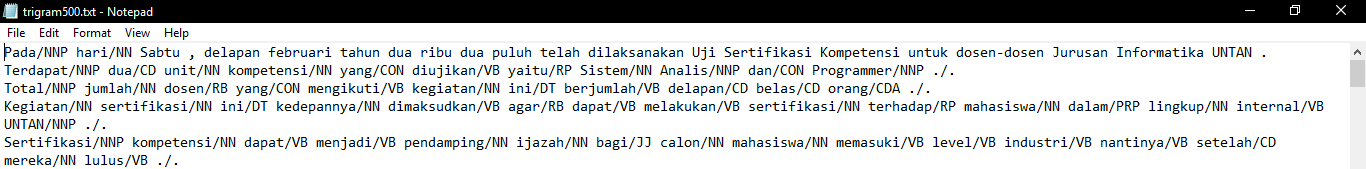
### 4.1.5 Hasil *Tagging* PoS Metode Trigram untuk Bahasa Indonesia

Pada metode *trigram* selanjutnya juga sama lakukan pengetikan koding pada aplikasi *phycharm community* kemudian jalankan dan akan menghasilkan hasil yang dapat di lihat pada gambar 4.5.

**Kode Program 4. 3** Proses Pelatihan Trigram

import nltk  
import trigram  
import time  
startTime = time.time()  
  
*#kalimat uji*kalimatuji=[]  
f1 = open("uji500.txt", "r")  
for baris in f1 :  
 kaluji = baris.strip()  
 kalimatuji.append(kaluji.split())  
 *#print(kalimatuji, len(kalimatuji))*train\_data=[]  
f0 = open("manual500.txt", "r") *#corpustrain.txt*for baris in f0 :  
 kal = baris.strip()  
 kalimat =[nltk.tag.str2tuple(t) for t in kal.split()]  
 *#print(kalimat)* train\_data.append(kalimat)  
  
*#print(train\_data)*datatrigram=[]  
trigram\_tagger = nltk.TrigramTagger(train\_data)  
for jumkal in range(len(kalimatuji)):  
 *#print("Unigram =", unigram\_tagger.tag(kalimatuji[jumkal]))* datatrigram.append([nltk.tag.tuple2str(x) for x in trigram\_tagger.tag(kalimatuji[jumkal])])  
  
  
x=0  
for x in range(0, len(datatrigram)):  
 y=0  
 txtOut = ""  
 for y in range(0, len(datatrigram[x])):  
 txtOut=txtOut + datatrigram[x][y]+ ' '  
 print(txtOut [:-1])  
  
  
*#for kalimat in datatrigram:  
# print(kalimat)  
#print(datatrigram[0])*"""oleh clik jawaban dulu lalu proses durasi --> stop time"""  
print(f"waktu eksekusi = {time.time()-startTime}")

Hasil dari metode trigram ini tentunya dapat kita langsung dengan dari gambar 4.5 yang dimana pada kalimat “ Pada/NNP hari/NN Sabtu , delapan februari tahun dua ribu dua puluh telah dilaksanakan Uji Sertifikasi Kompetensi untuk dosen-dosen Jurusan Informatika UNTAN .” pada teks kalimat ini dengan metode sebelumnya terdapat tagging namun pada kalimat ini tidak, dikarenakan setiap metode termasuk trigram memberikan nilai tagging yang manyatakan bahwa kalimat ini tidak ada nilai tagging, walaupun berdasarkan data latih yang sama dengan metode sebelumnya.



**Gambar 4. 5**Hasil Tagging Metode Trigram

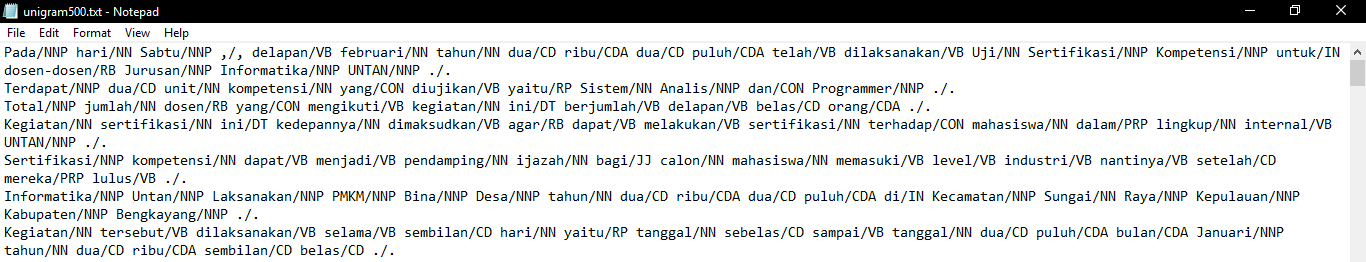
### Hasil *Tagging* PoS Metode Unigram untuk Bahasa Indonesia

Metode terakhir yang digunakan pada penitian ini adalah metode unigram yang dimana juga terdapat perbedaan dari kodingan dan hasil yang dihasilkan dari metode ini, dibawah ini adalah kodingan untuk proses tagging metode unigram.

**Kode Program 4. 4** Proses Pelatihan Unigram

import nltk  
import unigram  
import time  
startTime = time.time()  
  
*#kalimat uji*kalimatuji=[]  
f1 = open("uji500.txt", "r")  
for baris in f1 :  
 kaluji = baris.strip()  
 kalimatuji.append(kaluji.split())  
 *#print(kalimatuji, len(kalimatuji))*train\_data=[]  
f0 = open("manual500.txt", "r") *#corpustrain.txt*for baris in f0 :  
 kal = baris.strip()  
 kalimat =[nltk.tag.str2tuple(t) for t in kal.split()]  
 *#print(kalimat)* train\_data.append(kalimat)  
  
  
*#print(train\_data)*dataunigram=[]  
unigram\_tagger = nltk.UnigramTagger(train\_data)  
for jumkal in range(len(kalimatuji)):  
 *#print("Unigram =", unigram\_tagger.tag(kalimatuji[jumkal]))* dataunigram.append([nltk.tag.tuple2str(x) for x in unigram\_tagger.tag(kalimatuji[jumkal])])  
  
x=0  
for x in range(0, len(dataunigram)):  
 y=0  
 txtOut = ""  
 for y in range(0, len(dataunigram[x])):  
 txtOut=txtOut + dataunigram[x][y]+ ' '  
 print(txtOut [:-1])  
  
  
*#for kalimat in dataunigram:  
# print(kalimat)  
#print(dataunigram[0])*"""oleh clik jawaban dulu lalu proses durasi --> stop time"""  
print(f"waktu eksekusi = {time.time()-startTime}")

Setelah dilakukan proses tagging maka hasil yang didapatkan ada pada gambar 4.6 yang dimana pada gambar tersebut memiliki perbedaan kelas kata antara satu dan lainnya yang dimana pada metode trigram pada kalimat “ Pada/NNP hari/NN Sabtu/NNP ,/, delapan/VB februari/NN tahun/NN dua/CD ribu/CDA dua/CD puluh/CDA telah/VB dilaksanakan/VB Uji/NN Sertifikasi/NNP Kompetensi/NNP untuk/IN dosen-dosen/RB Jurusan/NNP Informatika/NNP UNTAN/NNP ./.” hanya terdapat beberapa kata saja yang terdapat tagging, namun pada metode unigram ini terdapat tagging bagi setiap kata dari kalimat yang contoh tersebut.



**Gambar 4. 6** Hasil Tagging Metode Unigram

* 1. **Pengujian**

#### 4.2.1.1 Hasil Pengujian Accuracy, Precision, Recall dan F-measure

Pengujian *accuracy, precision, recall,* dan *F-Measure* dilakukan dengan menggunakan dua buah korpus yang telah di latih sebelumnya. Setiap metode mempunyai korpus latih yang sama dan korpus uji yang sama tetapi memiliki hasil berbeda, karena setiap metode melakukan hasil *tagging* pada korpus uji dengan mendapatkan data dari korpus latih.

Setelah mendapatkan data tersebut dilakukan pengujian dengan cara menggunakan koding pada gambar 3.1 dan menggunakan tabel 3.1 untuk melakukan pendataan dari setiap nilai akurasi yang dihasilkan dari setiap metode.

**Tabel 4. 2** Tabel Hasil Pengujian Accuracy, Precision, Recall dan F-measure

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Metode PoS Tagger | Accuracy Score | Precision Score | Recall Score | F-Measure Score |
| Unigram | 0.966 | 1.0 | 0.950 | 0.974 |
| Bigram | 0.887 | 0.854 | 0.942 | 0.896 |
| Trigram | 0.889 | 0.860 | 0.952 | 0.904 |
| HMM | 0.966 | 0.977 | 0.966 | 0.971 |
| Perceptron | 0.988 | 0.979 | 0.978 | 0.978 |

Dari tabel 4.3 dapat dilihat dari tabel di atas untuk nilai *accuracy score* yang tertinggi adalah *Perceptron* yang dimana pada metode ini memiliki nilai yang lebih rendah pada pengujian *precision score, recall score, f-measure* sedangkan untuk nilai *precision score* tertinggi yaitu dari metode *Unigram* yang dimana metode ini juga memiliki nilai *accuracy* yang hanya berbeda 0,022 dari *Perception*, sedangkan untuk nilai *recall* yang tertinggi adalah metode *perceptron* yang memiliki nilai sebesar 0,978, dan pengujian terakhir yaitu *f-measure* adalah metode *perceptron* yaitu sebesar 0,978. Jika kita lihat menggunakan dengan grafik tentunya akan lebih mudah untuk di anilisa, sebagai berikut :

**Grafik 4. 1** Hasil Pengujian Metode PoS Tagger

Dapat dilihat pada grafik gambar 4 Nilai akurasi yang dihasilkan dari setiap metode PoS Tagger yang di uji memberikan nilai terhadap setiap pengujian yang dimana metode *unigram* memberikan nilai yang lebih tinggi daripada metode lainnya pada pengujian *precision* dengan nilai 1.0, sedangkan nilai *f-measure* tertinggi *Perceptron* sebesar 0,974 dan hanya berbeda 0,022 dari metode *perceptron* pada pengujian *accuracy*, sedangkan untuk pengujian *accuracy* metode *perceptron* memberikan nilai 0,988, dan untuk pengujian *recall* metode *Perceptron* memberikan nilai tertinggi yaitu sebesar 0,978.

* 1. **Analisa Hasil Pengujian**

Selanjutnya untuk hasil pengujian nilai akurasi dengan perhitungan kalimat 500 kalimat uji dan 500 kalimat latih yang memberikan nilai yang cukup tinggi yang dimana nilai dari setiap metode PoS Tagger berjumlah diatas 0.9. Tag Pos yang telah dikembangkan ini juga memberikan proses yang lebih cepat dalam melakukan perhitungan pengujian karena dengan jumlah tag PoS yang berjumlah 22 dapat mempercepat proses perhitungan dan juga mempercepat dalam proses pengerjaan dalam melakukan tagging karena dengan jumlah tag PoS yang lebih besar akan memberikan beban yang cukup besar dan bagi proses perhitungan dan pengujian yang dimana hal ini tentunya akan memperlambat proses penelitian.

# BAB V Kesimpulan dan Saran

## Kesimpulan

Berdasarkan hasil analisis dan pengujian terhadap nilai akurasi metode PoS Tagger untuk bahasa Indonesia, maka dapat ditarik kesimpulan sebagi berikut :

1. Hasil dari perbandingan tag PoS yang telah dilakukan menghasilkan tagging PoS baru yang berjumlah 22.
2. Nilai akurasi yang dihasil dari setiap metode PoS Tagger yang di uji memberikan nilai terhadap setiap pengujian yang dimana metode *Unigram* memberikan nilai yang lebih tinggi daripada metode lainnya pada pengujian *precision* dengan nilai 1,0, sedangkan nilai *F-measure* tertinggi yaitu *Perceptron* sebesar 0,978 dan hanya berbeda 0,022 dari metode *perceptron* pada pengujian *accuracy*, sedangkan untuk pengujian *accuracy* metode *perceptron* memberikan nilai 0,988, dan untuk pengujian *recall* metode *Perceptron* memberikan nilai tertinggi yaitu sebesar 0,978.

## Saran

Adapun beberapa hal yang perlu ditambahkan dalam pengembangan penelitian ini adalah sebagai berikut:

1. Perlu dilakukan pengembangan sistem *text to speech* bahasa Indonesia dengan manfaatkan tag PoS dari penelitian ini sehingga dapat dilakukan pembangkitan suara oleh sistem TTS.
2. Perlu dilakukan penambahan jumlah korpus dan dilakukan penelitian untuk mendapatkan nilai akurasi jeda terhadap kalimat berdasarkan tag PoS dari penelitian ini.

Daftar Pustaka

Akbar Khamsah. 2018. Pengaruh domain teks pada korpus terhadap akurasi mesin penerjemah statistik.

Abd Wahab Syahroni, Joan Santoso, Endang Setyati. (2017, Agustus 10). Pendekatan Rule Handmade untuk Menentukan Klausa Bahasa Indonesia. *Konferensi Nasional Sistem & Informatika 2017*.

Abney, S. P. (1991). *Parsing By Chunks, In Robert Berwick, Steven Abney, and Carol Tenny (eds)*. Kluwer Academic Publishers.

Adriani, M. 2009. *Developing Postag for Bahasa Indonesia*. PAN Localization Project*.* Universitas Indonesia.

Arry Akhmad Arman, A. B. (2013). *Syntatic Phrase Chunking for Indonesian Language*. Bandung: Institut Teknologi Bandung.

Atwell, C. B. (2007). Corpus-Based Evaluation of Prosodic Phrase Break Prediction Using nltk\_lite's Chunk Parser to Detect Prosodic Phrase Boundaries in the Aix-MARSEC Corpus of Spoken English. *Corpus* *Linguistics Conference 2007*. United Kingdom: University of Birmingham.

Christianti M, V.; Pragantha, J.; dan Victor. 2016. *Part-of-Speech Tagging untuk Bahasa Indonesia Menggunakan Stanford POS-Tagging.* Jakarta: Universitas Tarumanegara.

Do, Q. T., et al. 2015. *Improving Translation of Emphasis with Pause Prediction in Speech-to-speech Translation Systems*. Japan: Universitas Nagoya.

Fitriawati, Lia Suci. 2020. *Implementasi Text To Speech Pada Website Menggunakan Metode Shallow Parsing*. Pontianak: Universitas Tanjungpura.

Fitri, M. 2013. *Perancangan Sistem Temu Balik Informasi dengan Metode Pembobotan Kombinasi TF-IDF untuk Pencarian Dokumen Berbahasa Indonesia.* Pontianak: Universitas Tanjungpura.

Handel, V. R. 2008. *Uniform Observability of Hidden Markov Models and Filter Stability for Unstable Signals.* New Jersey: Universitas Princeton.

Kamiludin, M. I. 2017. *Prediksi Jeda pada Ucapan Bahasa Melayu Pontianak dengan Menggunakan Metode Shallow Parsing*. Pontianak: Universitas Tanjungpura.

Larasati, S. D.; Kuboň, V.; dan Zeman, D. 2011. *Indonesian Morphology Tool (MorphInd): Towards an Indonesian Corpus*. Springer CCIS proceedings of the Workshop on Systems and Frameworks for Computational Morphology*.* Zurich.

Li, J., et al. 2004. *Chinese Prosody Phrase Break Prediction Based on Maximum Entropy Model*. China: University of Science and Technology of China.

Lisangan, E. A. 2013. *Natural Language Processing dalam Memperoleh Informasi Akademik Mahasiswa Universitas Atma Jaya Makassar.* Makassar: Universitas Atma Jaya Makassar.

Litosseliti, L. 2010. *Research Methods in Linguistics.* New York: Continuum International Publishing Group.

Manurung, R. 2016. *Tutorial: Pengenalan terhadap POS tagging dan Probabilistic Parsing*. Workshop Nasional INACL*.* Universitas Indonesia.

Muniyati, Evi F. 2019. *Prediksi Jeda Dalam Ucapan Kalimat Bahasa Melayu Pontianak Menggunakan Hidden Markov Model Berbasis Part Of Speech*. Pontianak: Universitas Tanjungpura.

Magdalena, Y. 2019. *Prediksi Jeda pada Ucapan Kalimat Bahasa Melayu Pontianak Menggunakan Metode Shallow Parsing dengan Pengembangan Rule Grammar dan Rule Jeda.* Pontianak, Indonesia

Na'im, A. dan Syaputra, H. 2010. *Hasil Sensus Penduduk 2010: Kewarganegaraan, Suku Bangsa, Agama dan Bahasa Sehari-Hari Penduduk Indonesia.* Jakarta: Badan Pusat Statistik.

Nicolas, P. R. 2015. *Scala for Machine Learning*. Birmingham: Packt Publishing Ltd.

Novianti, E. 2011. *Menilik Nasib Bahasa Melayu Pontianak*. Seminar Internasional "Language Maintenance and Shift"*.* Universitas Diponegoro.

Nugraha, A. T. 2014. *Prediksi Jeda dalam Ucapan Kalimat Bahasa Indonesia dengan Hidden Markov Model.* Pontianak: Universitas Tanjungpura.

Paiki, F. F. 2006. *Evaluasi Penggunaan Similarity Thesaurus Terhadap Ekspansi Kueri pada Sistem Temu Kembali Informasi Berbahasa Indonesia.* Manokwari: Universitas Papua.

Pisceldo, F.; Adriani, M; dan Manurung, R. 2009. *Probabilistic Part Of Speech Tagging for Bahasa Indonesia*. Third International Wokshop on Malay and Indonesian Language Engineering*.* Singapore.

Prasetyo M. E. B. 2010. *Teori Dasar Hidden Markov Model*. Bandung: Institut Teknologi Bandung.

Rozi, I. F.; Pramono, S. H.; dan Dahlan, E. A. 2012. *Implementasi Opinion Mining (Analisis Sentimen) untuk Ekstraksi Data Opini pada Perguruan Tinggi.* Malang: Universitas Brawijaya.

Sarkar, P. dan Rao K., S. 2015. *Data-Driven Pause Prediction for Speech Synthesis in Storytelling Style Speech*. India: Indian Institute of Technology Kharagpur.

Setyaningsih, E. R. 2017. *Part of Speech Tagger untuk Bahasa Indonesia dengan Menggunakan Modifikasi Brill*. Surabaya: Sekolah Tinggi Teknik Surabaya.

Suciadi, J. 2001. *Studi Analisis Metode-Metode Parsing dan Interpretasi Semantik pada Natural Language Processing*. Jurnal Informatika.

WaveSurfer. www.speech.kth.se/wavesurfer/index.html. Diakses tanggal 28 Mei 2018.

Wibisono, Y. 2008. *Penggunaan Hidden Markov Model untuk Kompresi Kalimat.* Bandung: Institut Teknologi Bandung.

Wulandari, Y. 2020. *Aplikasi Pelaporan Bantuan Peternakan Di Kabupaten Sambas Dengan Metode Kepustakaan Dan Lapangan*. Pontianak : Universitas Tanjung Pura.

Wicaksono, A. F. dan Purwarianti, A. 2010. *HMM Based Part-of-Speech Tagger for Bahasa Indonesia*.Proceedings of 4th International MALINDO (Malay and Indonesian Language) Workshop*.*