**POS TAG BAHASA INDONESIA DENGAN TAGSET FLAIR**



**LAPORAN**  
**Disusun untuk Memenuhi Tugas Individu**  
**Mata Kuliah Pemrosesan Bahasa Alami Semester VIII**  
**yang diampu oleh Sukmawati Nur Endah, S.Si, M.Kom.**

**DISUSUN OLEH:**  
**LINGGAR MARETVA CENDANI (24060117120031)**

**PROGRAM STUDI STRATA 1 INFORMATIKA**  
**DEPARTEMEN ILMU KOMPUTER/INFORMATIKA**  
**FAKULTAS SAINS DAN MATEMATIKA**  
**UNIVERSITAS DIPONEGORO**  
**SEMARANG**  
**2021**

DAFTAR ISI

[DAFTAR ISI 2](#_Toc52999350)

[BAB I 3](#_Toc52999351)

[PENDAHULUAN 3](#_Toc52999352)

[1.1 Permasalahan 3](#_Toc52999353)

[1.2 Tujuan 3](#_Toc52999354)

[BAB II 4](#_Toc52999355)

[PEMBAHASAN 4](#_Toc52999356)

[2.1 POS Tagging dengan Flair 4](#_Toc52999357)

[2.2 Source Code Program](#_Toc52999358) 4

[a. Mount Google Drive 4](#_Toc52999357)

[b. Membaca File Berita 5](#_Toc52999357)

[c. Install Library Flair 5](#_Toc52999357)

[d. Import Semua Library Yang Akan Digunakan 5](#_Toc52999357)

[e. Me-load Korpus, Menentukan Tipe Tag, Membuat Tag Dictionary, Membuat Word Embedding, Membuat Tagger, dan Melakukan Training 5](#_Toc52999357)

[f. Me-load Tagger Yang Sudah Dibuat 6](#_Toc52999357)

[g. Memberikan POS Tag Pada Berita 1 6](#_Toc52999357)

[h. Memberikan POS Tag Pada Berita 2 6](#_Toc52999357)

[2.3 Output Program 7](#_Toc52999358)

[a. Hasil Training 7](#_Toc52999357)

[a. Berita 1 8](#_Toc52999357)

[a. Berita 2 9](#_Toc52999357)

[2.4 Analisa Hasil 10](#_Toc52999358)

[BAB III 1](#_Toc52999359)3

[PENUTUP 1](#_Toc52999360)3

[3.1 Kesimpulan 1](#_Toc52999361)3

BAB I

PENDAHULUAN

1. Permasalahan
2. Apa itu Flair dan bagaimana cara POS Tagging dengan Flair?
3. Bagaimana source code program untuk melakukan POS Tagging dengan Flair?
4. Bagaimana output dari program dalam melakukan POS Tagging dengan Flair?
5. Bagaimana analisa hasil dari proses POS Tagging pada artikel berita dengan Flair?
6. Tujuan
7. Mengetahui apa itu Flair dan cara melakukan POS Tagging dengan Flair
8. Mengetahui source code program untuk melakukan POS Tagging dengan Flair
9. Mengetahui output program dalam melakukan POS Tagging dengan Flair
10. Mengetahui analisa hasil dari proses POS Tagging pada artikel berita dengan Flair

BAB II

PEMBAHASAN

* 1. POS Tagging dengan Flair

***Word Clasess*** merupakan kelas kelas kata, biasa disebut juga ***Part of Speech (POS)***. Proses pemberian label pada tiap kata kelas kata yang sesuai disebut **POS Tagging**. Pada program ini, digunakan sebuah library Python yang bernama **Flair** untuk melakukan POS Tagging.

**Flair** sendiri merupakan library bahasa pemrograman Python yang khusus menangani kasus - kasus di bidang ***Natural Language Processing (NLP)***. Flair dapat digunakan untuk melakukan berbagai operasi NLP seperti *Word Embedding*, Tokenisasi, Klasifikasi Teks, dan tentu saja POS Tagging.

Flair memiliki koleksi korpus dari berbagai macam bahasa, termasuk Bahasa Indonesia. Korpus pada Falir sendiri terdiri dari 3 buah dataset korpus, yaitu *Development Corpus, Train Corpus, dan Test Corpus*. Maka dari itu, **Tagger** atau fungsi yang nantinya digunakan untuk melakukan POS Tagging dibuat dengan cara melakukan training pada corpus yang merupakan *tagset/ tag dictionary*.

Pada proses training, terdapat parameter yang dapat di-*tuning*, yaitu **learning\_rate** dan **max\_epoch.**

* 1. **Source Code Program**

1. **Mount Google Drive**

|  |
| --- |
| #mounting Google Drive  from google.colab import drive  drive.mount('/content/drive') |

1. **Membaca File Berita**

|  |
| --- |
| #read file  txt\_berita\_1 = open('/content/drive/MyDrive/Upload/Akademik/Tugas 2 PBA/Cetak Assist, Asnawi Mangkualam Masuk Best 11 K League 2.txt', 'r').read()  txt\_berita\_2 = open('/content/drive/MyDrive/Upload/Akademik/Tugas 2 PBA/Egy Maulana Vikri Sumbang Assist dan Gol untuk Lechia Gdansk II Sebelum Menghadap Shin Tae-yong.txt', 'r').read() |

1. **Install Library Flair**

|  |
| --- |
| !pip install flair |

1. **Import Semua Library Yang Akan Digunakan**

|  |
| --- |
| from flair.data\_fetcher import NLPTaskDataFetcher, NLPTask  from flair.embeddings import TokenEmbeddings, WordEmbeddings, StackedEmbeddings, BertEmbeddings  from typing import List |

1. **Me-load Korpus, Menentukan Tipe Tag, Membuat Tag Dictionary, Membuat Word Embedding, Membuat Tagger, dan Melakukan Training**

|  |
| --- |
| # 1. get the corpus  corpus = NLPTaskDataFetcher.load\_corpus(NLPTask.UD\_INDONESIAN)  # 2. what tag do we want to predict?  tag\_type = 'upos'  # 3. make the tag dictionary from the corpus  tag\_dictionary = corpus.make\_tag\_dictionary(tag\_type=tag\_type)  # 4. initialize embeddings  embedding\_types: List[TokenEmbeddings] = [      WordEmbeddings('id-crawl'),      WordEmbeddings('id'),      #WordEmbeddings('glove'),      #BertEmbeddings('bert-base-multilingual-cased')  ]  embeddings: StackedEmbeddings = StackedEmbeddings(embeddings=embedding\_types)  # 5. initialize sequence tagger  from flair.models import SequenceTagger  tagger: SequenceTagger = SequenceTagger(hidden\_size=256,                                          embeddings=embeddings,                                          tag\_dictionary=tag\_dictionary,                                          tag\_type=tag\_type,                                          use\_crf=True)  from flair.trainers import ModelTrainer  trainer: ModelTrainer = ModelTrainer(tagger, corpus)  # 7. start training  trainer.train('resources/taggers/example-universal-pos',                learning\_rate=0.1,                mini\_batch\_size=32,                max\_epochs=10) |

1. **Me-load Tagger Yang Sudah Dibuat**

|  |
| --- |
| from flair.data import Sentence  tag\_pos = SequenceTagger.load('resources/taggers/example-universal-pos/best-model.pt') |

1. **Memberikan POS Tag Pada Berita 1**

|  |
| --- |
| berita\_1 = Sentence(txt\_berita\_1)  tag\_pos.predict(berita\_1)  print(berita\_1.to\_tagged\_string()) |

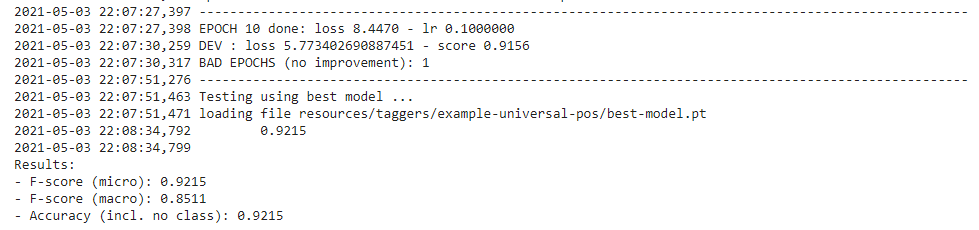
1. **Memberikan POS Tag Pada Berita 2**

|  |
| --- |
| berita\_2 = Sentence(txt\_berita\_2)  tag\_pos.predict(berita\_2)  print(berita\_2.to\_tagged\_string()) |

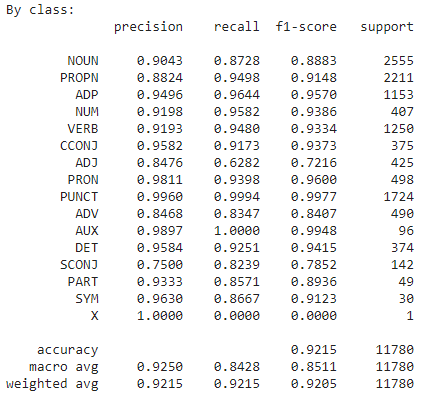
* 1. **Output Program**

1. **Hasil Training**

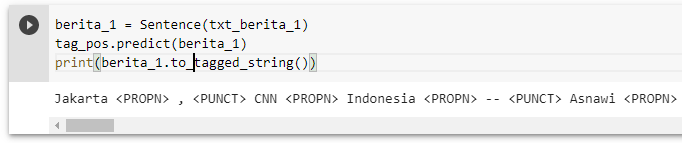
* **Overall Score :**



* **Score By Class :**

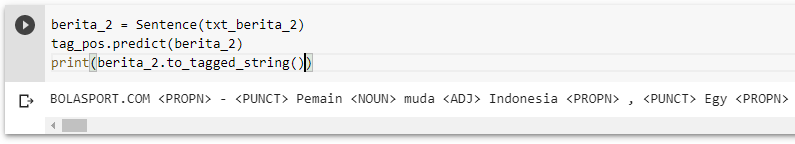


1. **Berita 1**



|  |
| --- |
| #**Cetak Assist, Asnawi Mangkualam Masuk Best 11 K League 2**  Jakarta <PROPN> , <PUNCT> CNN <PROPN> Indonesia <PROPN> -- <PUNCT> Asnawi <PROPN> Mangkualam <PROPN> masuk <VERB> dalam <ADP> daftar <NOUN> Best <PROPN> 11 <NUM> K <PROPN> League <PROPN> 2 <NUM> pekan <NOUN> kedelapan <NUM> setelah <ADP> membuat <VERB> assist <NOUN> kemenangan <NOUN> Ansan <PROPN> Greeners <PROPN> atas <ADP> Daejeon <PROPN> Citizens <PROPN> , <PUNCT> Sabtu <PROPN> ( <PUNCT> 24 <NUM> / <SYM> 4 <NUM> ) <PUNCT> . <PUNCT> Mantan <NOUN> pemain <NOUN> PSM <PROPN> Makassar <PROPN> itu <DET> dipilih <VERB> masuk <VERB> dalam <ADP> tim <NOUN> terbaik <ADJ> oleh <ADP> Federasi <PROPN> Sepak <PROPN> bola <NOUN> Profesional <PROPN> Korea <PROPN> pada <ADP> 27 <NUM> April <PROPN> . <PUNCT> Ini <DET> merupakan <VERB> kali <NOUN> pertama <NUM> Asnawi <PROPN> masuk <VERB> dalam <ADP> tim <NOUN> terbaik <ADJ> setelah <ADP> empat <NUM> kali <NOUN> memperkuat <VERB> Ansan <PROPN> di <ADP> laga <NOUN> K <PROPN> League <PROPN> 2 <NUM> . <PUNCT> Dalam <ADP> pertandingan <NOUN> akhir <NOUN> pekan <NOUN> lalu <ADP> tersebut <DET> , <PUNCT> Asnawi <PROPN> tampil <VERB> cukup <ADV> militan <ADJ> . <PUNCT> Ketika <SCONJ> menjadi <VERB> penyerang <NOUN> sayap <NOUN> kanan <ADJ> pada <ADP> babak <NOUN> pertama <NUM> , <PUNCT> Asnawi <PROPN> beberapa <DET> kali <NOUN> mengirim <VERB> umpan <NOUN> ke <ADP> kotak <NOUN> penalti <NOUN> . <PUNCT> Selain <ADP> itu <DET> Asnawi <PROPN> juga <ADV> cukup <ADV> aktif <VERB> membantu <VERB> pertahanan <NOUN> . <PUNCT> Sempat <ADV> gagal <VERB> mencetak <VERB> gol <NOUN> ketika <SCONJ> mendapat <VERB> peluang <NOUN> yang <PRON> sangat <ADV> terbuka <ADJ> di <ADP> menit-menit <NOUN> akhir <NOUN> babak <NOUN> pertama <NUM> , <PUNCT> Asnawi <PROPN> yang <PRON> menjadi <VERB> bek <NOUN> kanan <ADJ> pada <ADP> babak <NOUN> kedua <NUM> kemudian <ADV> membuat <VERB> assist <NOUN> pada <ADP> 10 <NUM> menit <NOUN> akhir <NOUN> pertandingan <NOUN> . <PUNCT> Umpan <PROPN> Asnawi <PROPN> yang <PRON> dituntaskan <VERB> Shim <PROPN> Jae <PROPN> Min <PROPN> menjadi <VERB> gol <NOUN> kemenangan <NOUN> Ansan <PROPN> atas <ADP> Daejeon <PROPN> . <PUNCT> Sebelumnya <ADV> Ansan <PROPN> tidak <PART> pernah <ADV> menang <VERB> atas <ADP> Daejeon <PROPN> sejak <ADP> pertengahan <NOUN> 2019 <NUM> . <PUNCT> Ansan <PROPN> sempat <ADV> menempati <VERB> peringkat <NOUN> kedua <NUM> setelah <ADP> kemenangan <NOUN> atas <ADP> Daejeon <PROPN> , <PUNCT> namun <SCONJ> kini <ADV> klub <NOUN> berjuluk <VERB> Green <PROPN> Wolf <PROPN> itu <DET> merosot <VERB> ke <ADP> peringkat <NOUN> keempat <NUM> menyusul <ADP> tiga <NUM> poin <NOUN> yang <PRON> didapat <VERB> Jeonnam <PROPN> Dragons <PROPN> dan <CCONJ> FC <PROPN> Anyang <PROPN> . <PUNCT> Pada <ADP> pekan <NOUN> mendatang <VERB> , <PUNCT> Ansan <PROPN> akan <ADV> berhadapan <VERB> dengan <ADP> Seoul <PROPN> E-Land <PROPN> yang <PRON> menempati <VERB> peringkat <NOUN> kelima <NUM> . <PUNCT> Asnawi <PROPN> pun <ADV> berpeluang <VERB> kembali <ADV> dimainkan <VERB> , <PUNCT> namun <SCONJ> di <ADP> sisi <NOUN> lain <ADJ> pemain <NOUN> 21 <NUM> tahun <NOUN> itu <DET> juga <ADV> termasuk <VERB> dalam <ADP> pemain <NOUN> yang <PRON> dipanggil <VERB> ke <ADP> dalam <ADP> pemusatan <NOUN> latihan <NOUN> Timnas <PROPN> Indonesia <PROPN> jelang <PROPN> Kualifikasi <PROPN> Piala <PROPN> Dunia <PROPN> 2022 <NUM> . <PUNCT> Pelatnas <PROPN> akan <ADV> dimulai <VERB> pada <ADP> 1 <NUM> Mei <PROPN> hingga <ADP> 10 <NUM> Mei <PROPN> di <ADP> Jakarta <PROPN> , <PUNCT> setelah <ADP> itu <DET> pelatih <NOUN> Shin <PROPN> Tae <PROPN> Yong <PROPN> akan <ADV> membawa <VERB> anak <NOUN> asuhnya <NOUN> berlatih <VERB> di <ADP> Uni <PROPN> Emirat <PROPN> Arab <PROPN> sebelum <ADP> menghadapi <VERB> tiga <NUM> laga <NOUN> sisa <NOUN> Kualifikasi <PROPN> Piala <PROPN> Dunia <PROPN> pada <ADP> Juni <PROPN> . <PUNCT> Beberapa <DET> pekan <NOUN> lalu <CCONJ> Shin <PROPN> Tae <PROPN> Yong <PROPN> sudah <ADV> bertemu <VERB> dengan <ADP> perwakilan <NOUN> Ansan <PROPN> termasuk <VERB> membicarakan <VERB> perihal <ADP> pemanggilan <NOUN> Asnawi <PROPN> ke <ADP> Timnas <PROPN> Indonesia <PROPN> . <PUNCT> |

1. **Berita 2**



|  |
| --- |
| #**Egy Maulana Vikri Sumbang Assist dan Gol untuk Lechia Gdansk II Sebelum Menghadap Shin Tae-yong**  Jakarta <PROPN> , <PUNCT> CNN <PROPN> Indonesia <PROPN> -- <PUNCT> Asnawi <PROPN> Mangkualam <PROPN> masuk <VERB> dalam <ADP> daftar <NOUN> Best <PROPN> 11 <NUM> K <PROPN> League <PROPN> 2 <NUM> pekan <NOUN> kedelapan <NUM> setelah <ADP> membuat <VERB> assist <NOUN> kemenangan <NOUN> Ansan <PROPN> Greeners <PROPN> atas <ADP> Daejeon <PROPN> Citizens <PROPN> , <PUNCT> Sabtu <PROPN> ( <PUNCT> 24 <NUM> / <SYM> 4 <NUM> ) <PUNCT> . <PUNCT> Mantan <NOUN> pemain <NOUN> PSM <PROPN> Makassar <PROPN> itu <DET> dipilih <VERB> masuk <VERB> dalam <ADP> tim <NOUN> terbaik <ADJ> oleh <ADP> Federasi <PROPN> Sepak <PROPN> bola <NOUN> Profesional <PROPN> Korea <PROPN> pada <ADP> 27 <NUM> April <PROPN> . <PUNCT> Ini <DET> merupakan <VERB> kali <NOUN> pertama <NUM> Asnawi <PROPN> masuk <VERB> dalam <ADP> tim <NOUN> terbaik <ADJ> setelah <ADP> empat <NUM> kali <NOUN> memperkuat <VERB> Ansan <PROPN> di <ADP> laga <NOUN> K <PROPN> League <PROPN> 2 <NUM> . <PUNCT> Dalam <ADP> pertandingan <NOUN> akhir <NOUN> pekan <NOUN> lalu <ADP> tersebut <DET> , <PUNCT> Asnawi <PROPN> tampil <VERB> cukup <ADV> militan <ADJ> . <PUNCT> Ketika <SCONJ> menjadi <VERB> penyerang <NOUN> sayap <NOUN> kanan <ADJ> pada <ADP> babak <NOUN> pertama <NUM> , <PUNCT> Asnawi <PROPN> beberapa <DET> kali <NOUN> mengirim <VERB> umpan <NOUN> ke <ADP> kotak <NOUN> penalti <NOUN> . <PUNCT> Selain <ADP> itu <DET> Asnawi <PROPN> juga <ADV> cukup <ADV> aktif <VERB> membantu <VERB> pertahanan <NOUN> . <PUNCT> Sempat <ADV> gagal <VERB> mencetak <VERB> gol <NOUN> ketika <SCONJ> mendapat <VERB> peluang <NOUN> yang <PRON> sangat <ADV> terbuka <ADJ> di <ADP> menit-menit <NOUN> akhir <NOUN> babak <NOUN> pertama <NUM> , <PUNCT> Asnawi <PROPN> yang <PRON> menjadi <VERB> bek <NOUN> kanan <ADJ> pada <ADP> babak <NOUN> kedua <NUM> kemudian <ADV> membuat <VERB> assist <NOUN> pada <ADP> 10 <NUM> menit <NOUN> akhir <NOUN> pertandingan <NOUN> . <PUNCT> Umpan <PROPN> Asnawi <PROPN> yang <PRON> dituntaskan <VERB> Shim <PROPN> Jae <PROPN> Min <PROPN> menjadi <VERB> gol <NOUN> kemenangan <NOUN> Ansan <PROPN> atas <ADP> Daejeon <PROPN> . <PUNCT> Sebelumnya <ADV> Ansan <PROPN> tidak <PART> pernah <ADV> menang <VERB> atas <ADP> Daejeon <PROPN> sejak <ADP> pertengahan <NOUN> 2019 <NUM> . <PUNCT> Ansan <PROPN> sempat <ADV> menempati <VERB> peringkat <NOUN> kedua <NUM> setelah <ADP> kemenangan <NOUN> atas <ADP> Daejeon <PROPN> , <PUNCT> namun <SCONJ> kini <ADV> klub <NOUN> berjuluk <VERB> Green <PROPN> Wolf <PROPN> itu <DET> merosot <VERB> ke <ADP> peringkat <NOUN> keempat <NUM> menyusul <ADP> tiga <NUM> poin <NOUN> yang <PRON> didapat <VERB> Jeonnam <PROPN> Dragons <PROPN> dan <CCONJ> FC <PROPN> Anyang <PROPN> . <PUNCT> Pada <ADP> pekan <NOUN> mendatang <VERB> , <PUNCT> Ansan <PROPN> akan <ADV> berhadapan <VERB> dengan <ADP> Seoul <PROPN> E-Land <PROPN> yang <PRON> menempati <VERB> peringkat <NOUN> kelima <NUM> . <PUNCT> Asnawi <PROPN> pun <ADV> berpeluang <VERB> kembali <ADV> dimainkan <VERB> , <PUNCT> namun <SCONJ> di <ADP> sisi <NOUN> lain <ADJ> pemain <NOUN> 21 <NUM> tahun <NOUN> itu <DET> juga <ADV> termasuk <VERB> dalam <ADP> pemain <NOUN> yang <PRON> dipanggil <VERB> ke <ADP> dalam <ADP> pemusatan <NOUN> latihan <NOUN> Timnas <PROPN> Indonesia <PROPN> jelang <PROPN> Kualifikasi <PROPN> Piala <PROPN> Dunia <PROPN> 2022 <NUM> . <PUNCT> Pelatnas <PROPN> akan <ADV> dimulai <VERB> pada <ADP> 1 <NUM> Mei <PROPN> hingga <ADP> 10 <NUM> Mei <PROPN> di <ADP> Jakarta <PROPN> , <PUNCT> setelah <ADP> itu <DET> pelatih <NOUN> Shin <PROPN> Tae <PROPN> Yong <PROPN> akan <ADV> membawa <VERB> anak <NOUN> asuhnya <NOUN> berlatih <VERB> di <ADP> Uni <PROPN> Emirat <PROPN> Arab <PROPN> sebelum <ADP> menghadapi <VERB> tiga <NUM> laga <NOUN> sisa <NOUN> Kualifikasi <PROPN> Piala <PROPN> Dunia <PROPN> pada <ADP> Juni <PROPN> . <PUNCT> Beberapa <DET> pekan <NOUN> lalu <CCONJ> Shin <PROPN> Tae <PROPN> Yong <PROPN> sudah <ADV> bertemu <VERB> dengan <ADP> perwakilan <NOUN> Ansan <PROPN> termasuk <VERB> membicarakan <VERB> perihal <ADP> pemanggilan <NOUN> Asnawi <PROPN> ke <ADP> Timnas <PROPN> Indonesia <PROPN> . <PUNCT> |

* 1. Analisa Hasil

Pada program ini, digunakan 2 buah berita yang memiliki tema Sepakbola, dengan berita 1 yang berjudul "**Cetak Assist, Asnawi Mangkualam Masuk Best 11 K League 2**", dan berita 2 yang berjudul "**Egy Maulana Vikri Sumbang Assist dan Gol untuk Lechia Gdansk II Sebelum Menghadap Shin Tae-yong**". Kedua berita berformat txt, dan di load dari Google Drive.

Korpus yang digunakan adalah korpus Bahasa Indonesia, yang diload dengan memanggil **NLPTask.UD\_INDONESIAN**, yang dimana terdapat 3 set data korpus yang tadi telah dijelaskan (*development, train, dan test corpus*). Digunakan jenis tag-nya adalah **'upos'** karena akan dilakukan POS Tagging. Tag Dictionary dibuat dari korpus yang tadi telah diload, lalu Word Embedding yang digunakan adalah **'id-crawl' dan 'id'.** Pada proses training, digunakan parameter **learning\_rate dengan nilai 0,1 dan max\_epochs-nya 10**.

Proses training yang dilakukan pada korpus dengan jumlah data train 4477, development 559, dan test 557 kalimat didapatkan hasil terbaik yaitu besar **loss sebesar 5.773402690887451** dan **score-nya mencapai 0.9156**. Sedangkan F-score dan Akurasinya adalah sebagai berikut :

* F-score (micro): 0.9215
* F-score (macro): 0.8511
* Accuracy (incl. no class): 0.9215

Karena digunakan POS Tagging Universal, maka kelas kata yang kita miliki adalah sebagai berikut :

* ADJ : kata sifat
* ADP : preposisi
* ADV : keterangan
* AUX : kata bantu
* CCONJ : kata penghubung
* SCONJ : subordinating conjunctions
* DET : determiner
* NOUN : kata benda
* NUM : angka
* PART : partikel
* PROPN : proper noun (nama diri)
* PRON : kata ganti
* PUNCT : tanda baca
* SYM : simbol
* VERB : kata kerja
* X : lainnya

Dari hasil POS Tagging yang dilakukan pada kedua berita, menunjukkan hasil yang cukup bagus dan akurat. Karena kelas kata yang digunakan adalah universal, maka tiap kata diberi label kelas sesuai kelas universal yang sesuai.

Namun, karena hanya melakukan operasi POS Tagging dan belum melakukan operasi Name Entity Recognition, beberapa entitas yang memiliki lebih dari satu kata dan merupakan kesatuan diberikan label kelas secara terpisah, meskipun labelnya sama, namun pemberian label dilakukan per kata, bukan per entitas.

Namun secara keseluruhan hasil dari pemberian label kelas kata pada tiap kata telah sesuai dan dapat dikatakan sangat baik .

BAB III

PENUTUP

3.1 Kesimpulan

POS Tagging merupakan proses pemberian label pada tiap kata kelas kata yang sesuai disebut. Flair yang merupakan library Python yang dapat digunakan untuk melakukan POS Tagging memberikan hasil yang memuaskan.

Untuk proses tag POS pada dokumen Bahasa Indonesia, didapatkan hasil akurasi mencapai 92,15 %. Dan hasil pemberian tag POS pada dokumen berita memberikan hasil yang baik dan tiap kata diberi label sesuai jenis kelas katanya yang pada program ini digunakan kelas kata universal.