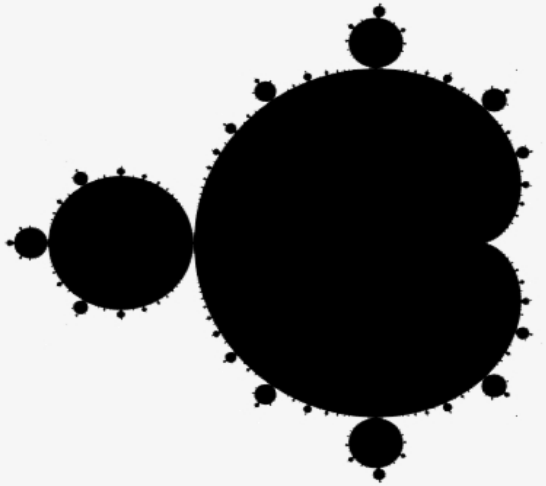
TextBlob

An Intuitive Interface for NLTK

Presented by
Rajashri Chitti



TextBlob

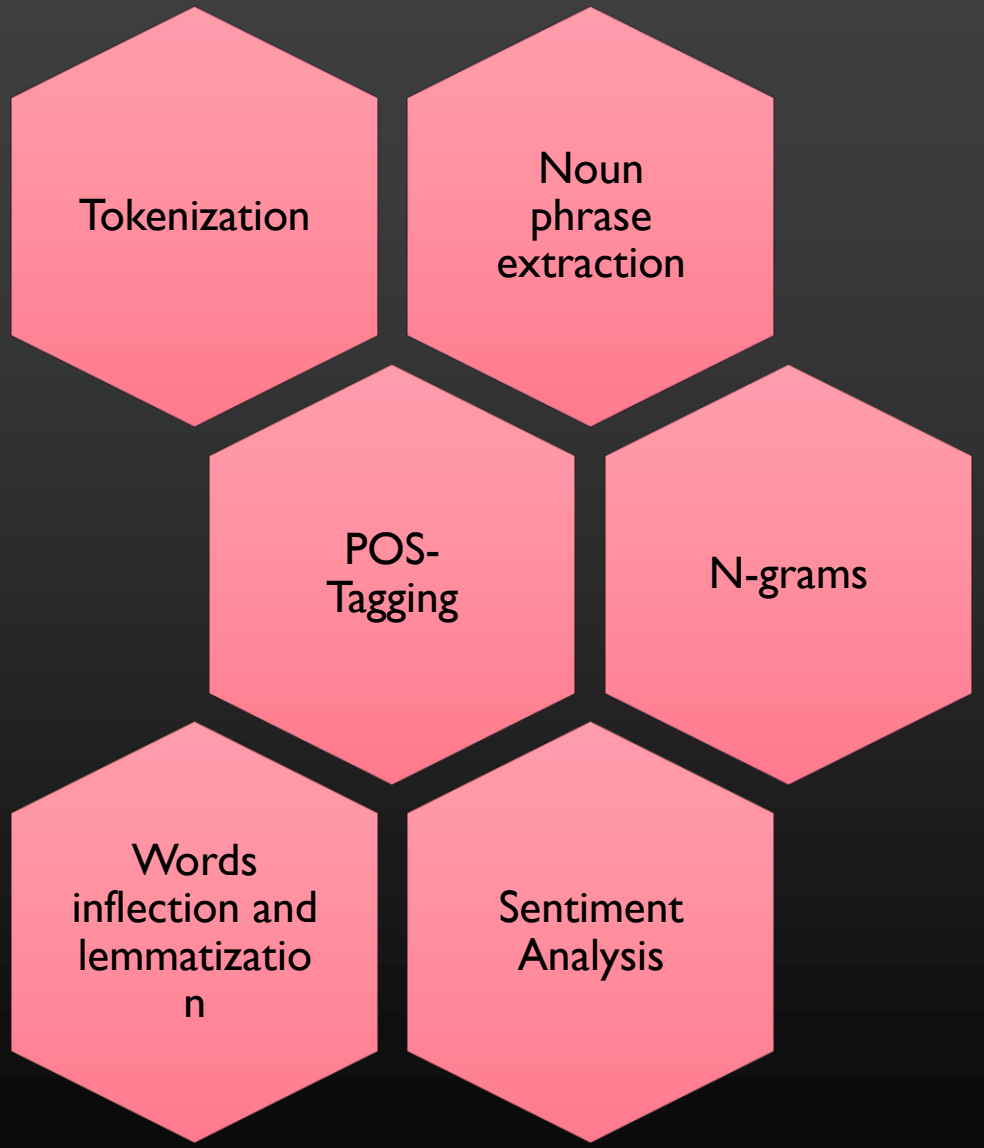
Contents

- ✓Introduction to TextBlob
- ✓Tasks handled by TextBlob
- ✓Other tasks
- ✓Pros and cons
- ✓Conclusion
- ✓References

Introduction to TextBlob

- ✓ TextBlob is a python library for processing text-based information.
- ✓ It gives a basic API to plunging into normal characteristic language preparing (NLP) tasks.
- ✓ Such as grammatical feature labelling, noun phrase extraction, sentiment analysis, classification, translation, and many more.
- ✓ Setting up the system
  - pip install –U textblob
  - python –m textblob.download_corpora

Tasks handled by TextBlob

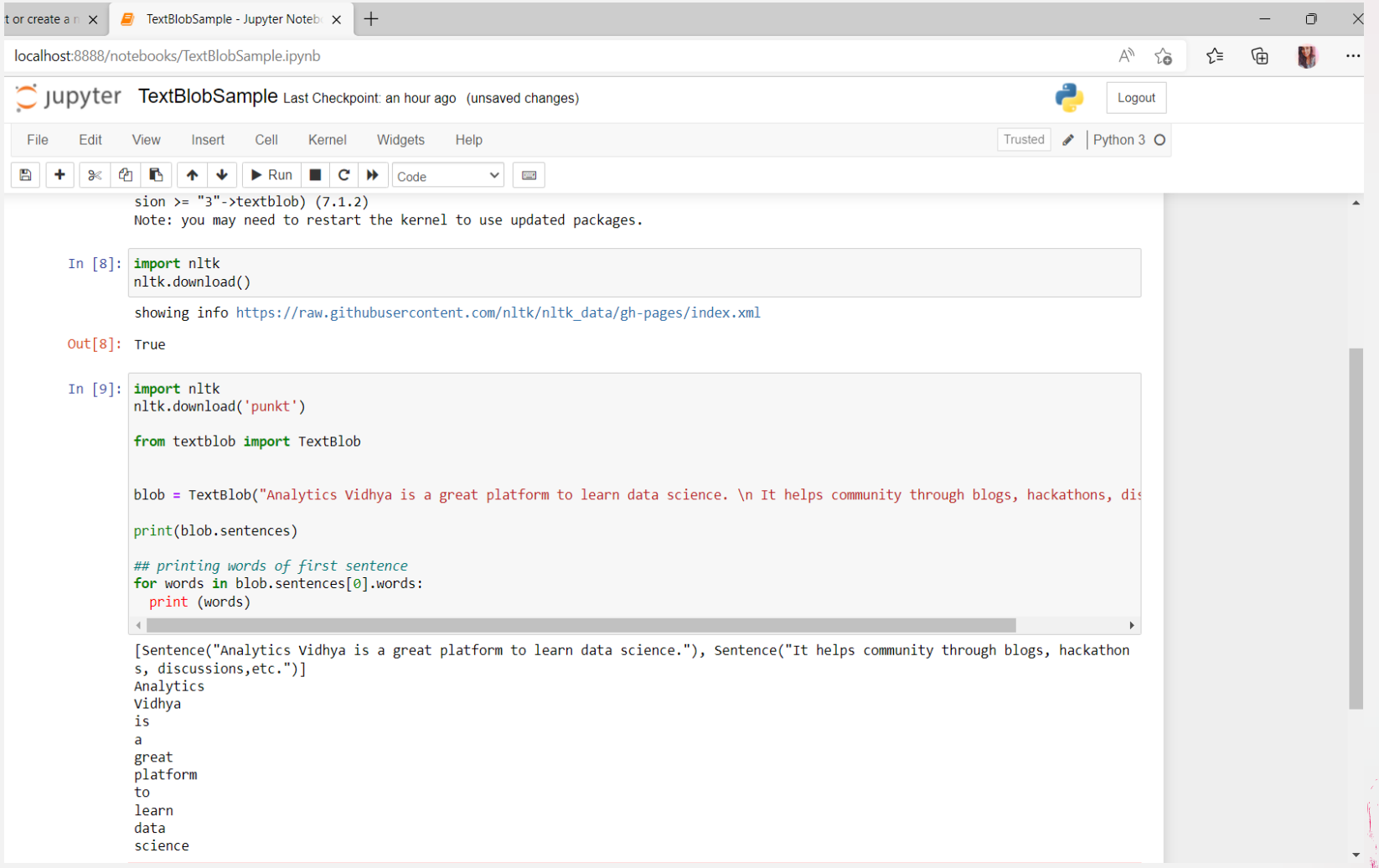


1. Tokenization

To do this using TextBlob, follow the two steps:

Create a **textblob** object and pass a string with it.

Call **functions** of textblob in order to do a specific task.



2. Noun Phrase Extraction

Noun phrase extraction **takes part of speech type into account when determining relevance**.

```
In [10]: blob = TextBlob("Analytics Vidhya is a great platform to learn data science.")
         for np in blob.noun_phrases:
          print (np)

         analytics vidhya
         great platform
         data science
```

3. Part-of-speech Tagging

Part-of-speech tagging or grammatical tagging is a method to mark words present in a text on the basis of its definition and context.

```
In [12]: for words, tag in blob.tags:
          print (words, tag)

         Analytics NNS
         Vidhya NNP
         is VBZ
         a DT
         great JJ
         platform NN
         to TO
         learn VB
         data NNS
         science NN
```

4. Word Inflection & Lemmatization

Inflection is a process of *word* formation in which characters are added to the base form of a *word* to express grammatical meanings.

```
In [13]: blob = TextBlob("Analytics Vidhya is a great platform to learn data science. \n It helps community through blogs, hackathons, dis
         print (blob.sentences[1].words[1])
         print (blob.sentences[1].words[1].singularize())

         helps
         help
```

```
In [14]: from textblob import Word
         w = Word('Platform')
         w.pluralize()

Out[14]: 'Platforms'
```

```
In [17]: ## using tags
         for word,pos in blob.tags:
             if pos == 'NN':
                 print(word.pluralize())

         platforms
         sciences
         communities
```

```
In [18]: ## lemmatization
         w = Word('running')
         w.lemmatize("v") ## v here represents verb

Out[18]: 'run'
```

5. N-grams

- A combination of multiple words together are called N-Grams.
- N grams (N > 1) are generally more informative as compared to words, and can be used as features for language modelling.

```
In [20]: for ngram in blob.ngrams(2):
             print(ngram)
```

```
['Analytics', 'Vidhya']
['Vidhya', 'is']
['is', 'a']
['a', 'great']
['great', 'platform']
['platform', 'to']
['to', 'learn']
['learn', 'data']
['data', 'science']
['science', 'It']
['It', 'helps']
['helps', 'community']
['community', 'through']
['through', 'blogs']
['blogs', 'hackathons']
['hackathons', 'discussions']
['discussions', 'etc']
```

6. Sentiment Analysis

The sentiment function of textblob returns two properties, **polarity**, and **subjectivity.**

- Polarity is float which lies in the range of [-1,1] where 1 means positive statement and -1 means a negative statement.
- Subjective sentences generally refer to personal opinion, emotion or judgment whereas objective refers to factual information.

```
In [21]: print (blob)
         blob.sentiment

         Analytics Vidhya is a great platform to learn data science.
          It helps community through blogs, hackathons, discussions,etc.

Out[21]: Sentiment(polarity=0.8, subjectivity=0.75)
```

Other tasks

Spelling correction

```
In [23]: blob = TextBlob('I havv goood speling!')
         blob.correct()

Out[23]: TextBlob("I have good spelling!")
```

Word count

```
In [41]: blog = TextBlob('We are no longer the Knights who say Ni ekki. We are now the Knights who say Ekki ekki ekki PTANG.')
         blog.word_counts['ekki']

Out[41]: 4
```

Classification model like Naïve Bayes, etc.

Pros and Cons of TextBlob

Pros

- Since, it is built on the shoulders of NLTK and Pattern, therefore making it simple for beginners by providing an intuitive interface to NLTK.
- It provides language translation and detection which is powered by Google Translate (not provided with Spacy).

Cons

- It is little slower in the comparison to spacy but faster than NLTK. (Spacy > TextBlob > NLTK)
- It does not provide features like neural network model, integrated word vectors etc. which is provided by spacy.

Conclusion

- ✓ TextBlob, actually provided a very easy interface for beginners to learn basic NLP tasks.
- ✓ Textblob provides a wide variety of functions that are used to draw certain properties of the textual data.
- ✓ It allows us to change the properties of data to make it useful to pass it to the machine learning model.
- ✓ It is also useful in creating short summary of text, translation and language detection, and text classification is possible using this library.

References

- https://textblob.readthedocs.io/en/dev/
- https://www.analyticsvidhya.com/blog/2021/10/making-natural-language-processing-easy-with-textblob/
- https://www.topcoder.com/thrive/articles/getting-started-with-textblob-for-sentiment-analysis
- https://github.com/sloria/TextBlob
- https://neptune.ai/blog/sentiment-analysis-python-textblob-vs-vader-vs-flair

Thank You