Fake New Detection Using ML and NLP

Week 1:

**IDE:**

**Google Colab:** Click this link <https://colab.research.google.com/>

And follow this video [Google Colab Tutorial for Beginners | Get Started with Google Colab](https://youtu.be/RLYoEyIHL6A?si=y9T9JQJTPpfr3lFe)

We will be using google colab for writing all the codes.

**1. Python Basics:**

* By python we mean to learn intermediate-advanced python which will be necessary for the project. You may learn intermediate Python from here:  
  <https://www.learnpython.dev/01-introduction/>
* For deep learning techniques, we would be learning Advanced Python

<https://github.com/PacktPublishing/Advanced-Deep-Learning-with-Python>

* Go through this tutorial from Python Syntax to Python Arrays and keep practising by writing the codes in google colab. This is for absolute beginners.

<https://www.w3schools.com/python/>

[Learn Python - Full Course for Beginners [Tutorial]](https://www.youtube.com/watch?v=rfscVS0vtbw)

* Slides for your reference:

[Advance\_Python.pdf](https://drive.google.com/file/d/11LrkwmPq-rpGPY-HdmXhgXyGdGrs4PYR/view?usp=drive_link)

**Numpy**

* Numpy is a python library that stands for ‘numerical python’.
* While python allows working with lists containing elements of different data types, numpy is used for the manipulation of lists containing elements of one data type only.
* It adds powerful data structures to Python that guarantee efficient calculations with arrays and matrices and it supplies an enormous library of high-level mathematical functions that operate on these arrays and matrices.

You’ll find the following resources helpful -

* The best resource to turn to will always be the documentation - <https://numpy.org/doc/stable/user/basics.html>
* This is a fun tutorial-

<https://numpy.org/doc/stable/user/absolute_beginners.html#:~:text=One%20way%20we%20can%20initialize,%2D%20or%20higher%2Ddimensional%20data.&text=We%20can%20access%20the%20elements,in%20NumPy%20starts%20at%200>.

* You can also study the usage here - <https://www.geeksforgeeks.org/introduction-to-numpy/?ref=lbp>

**Pandas**

* Pandas is a python library that is used for data analysis and manipulation.

You’ll find the following resources helpful-

* Introduction to pandas - <https://www.nvidia.com/en-us/glossary/pandas-python/>
* Documentation :) - <https://pandas.pydata.org/docs/>
* Some tutorials to make life easier - <https://pandas.pydata.org/docs/getting_started/intro_tutorials/index.html>
* Study the usage - <https://www.geeksforgeeks.org/introduction-to-pandas-in-python/>

**Matplotlib**

* It is a python library used for visualisation of data made by John D. Hunter.
* Fun fact- Hunter initially developed Matplotlib during his postdoctoral research in neurobiology to visualize electrocorticography (ECoG) data of epilepsy patients.

You’ll find the following resources useful -

* I’m happy I stumbled across this article while compiling resources for you - <https://medium.com/mlpoint/matplotlib-for-machine-learning-6b5fcd4fbbc7>
* Documentation and tutorial - <https://matplotlib.org/stable/index.html>
* :) - <https://matplotlib.org/cheatsheets/>
* How to use it - <https://www.w3schools.com/python/matplotlib_intro.asp>

Seaborn

* Seaborn is a library built on top of matplotlib for visualising.

You’ll find the following resources helpful-

* Why do I need to study two of them? - <https://techifysolutions.com/blog/seaborn-visualize-data-beyond-matplotlib/#:~:text=Matplotlib%20lets%20users%20customize%20the,for%20certain%20type%20of%20plots>.
* These are all the plots - <https://seaborn.pydata.org/examples/index.html>
* Lovely read - <https://machinelearningmastery.com/seaborn-data-visualization-for-machine-learning/>

**Machine learning basics:**

**Machine Learning playlist (youtube) -** [**https://youtube.com/playlist?list=PLxCzCOWd7aiEXg5BV10k9THtjnS48yI-T&si=veUMmfbENMqXd5Nb**](https://youtube.com/playlist?list=PLxCzCOWd7aiEXg5BV10k9THtjnS48yI-T&si=veUMmfbENMqXd5Nb)

**w3schools tutorials link:** [**https://www.w3schools.com/python/python\_ml\_getting\_started.asp**](https://www.w3schools.com/python/python_ml_getting_started.asp)

you may start from linear regression in the above tutorial

**Natural Language Processing (NLP) foundations:**

The youtube playlist below is very easy to understand and beginner-friendly introduction to NLP. You may watch it when we start with NLP

[**https://youtube.com/playlist?list=PLZoTAELRMXVNNrHSKv36Lr3\_156yCo6Nn&si=L-T\_K7AtcDCO2MCO**](https://youtube.com/playlist?list=PLZoTAELRMXVNNrHSKv36Lr3_156yCo6Nn&si=L-T_K7AtcDCO2MCO)

If you want to with the text :[Natural Language Processing (NLP) Tutorial |](https://www.geeksforgeeks.org/natural-language-processing-nlp-tutorial/)

WEEK 2:

**Resources:**

**PPT LINK:** [**Natural Language Processing**](https://docs.google.com/presentation/d/1RltrhXev_8eDtUkGSTA-H5zCu3G6Tn63/edit?usp=drive_link&ouid=114882704169705368915&rtpof=true&sd=true)

* 1. EDA :
     1. <https://www.geeksforgeeks.org/what-is-exploratory-data-analysis/>
     2. [Exploratory Data Analysis (EDA) Using Python | Python Data Analysis | Python Training | Edureka](https://www.youtube.com/watch?v=-o3AxdVcUtQ)
  2. Pre-Processing :
     1. [Tokenization](https://youtu.be/fNxaJsNG3-s?si=XRT_w6F8Fk7YMcJu)
     2. https://www.geeksforgeeks.org/what-is-tokenization/
     3. Stemming and Lemmatization:
        1. <https://www.ibm.com/topics/stemming-lemmatization>
        2. [Stemming and Lemmatization: NLP Tutorial For Beginners - S1 E10](https://www.youtube.com/watch?v=HHAilAC3cXw)
  3. Feature Extraction :
     1. <https://www.geeksforgeeks.org/ml-one-hot-encoding/>
     2. <https://neptune.ai/blog/vectorization-techniques-in-nlp-guide>

4. Model Selection :

* 1. <https://towardsdatascience.com/top-machine-learning-algorithms-for-classification-2197870ff501>

5. Evaluation:

* 1. <https://www.analyticsvidhya.com/blog/2021/07/metrics-to-evaluate-your-classification-model-to-take-the-right-decisions/>

6. Hyper Parameter Tuning: (i) [Hyperparameter Tuning](https://www.geeksforgeeks.org/hyperparameter-tuning/)