Fake News Detection Using Machine Learning Notes

1. patch\_sklearn() in method in sklearnex from the intel oneApi toolkit and its was another modified version of the sklearn package, Which help to make the code to run 100x faster than before.

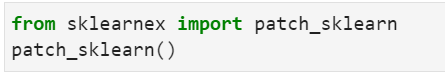


Fig 1.1

1. word\_tokenize ()

1. We use the method of word\_tokenize() to split sentence into words.

2. Is helpful to convert the text:string into numeric values to train the model.

1. Lemmatization ()

1.Instead of using the Stemming we can also use the lemmatization which provide the better performance output and it also perform based on morphological analysis.

2.Example ['Alice' , 'Studies' , 'Studying'] --> lemmatizer process -> output : ['Alice' , 'Studies' , 'Study']

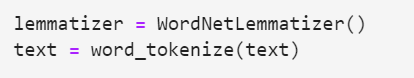


Fig 1.2

1. CountVectorizer ()

Is a text vectorization method provided by the scikit-learn library in Python. It is a simple and commonly used technique that converts a collection of text documents into a matrix of token counts.

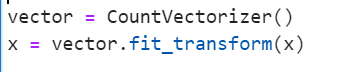


Fig 1.3

1. train\_test\_split ()

function is used to separate the (x,y) : dependent and independent values to train and test data in which the train variable are used to train the dataset and test datasets are used to test and find the prediction of model.

1. Tokenizer ()

1) keras.preprocessing.text.Tokenizer is a very useful tokenizer for text processing.

2) Tokenizer assumes that the word tokens of the input texts have been delimited by whitespaces.

3) Tokenizer provides the following functions:

1) It will first create a dictionary for the entire corpus (a mapping of each word token and its unique integer index index) (Tokenizer.fit\_on\_text())

2) It can then use the corpus dictionary to convert words in each corpus text into integer sequences Tokenizer.texts\_to\_sequences()

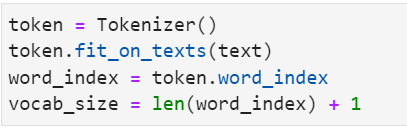


Fig 1.4

1. pad\_sequences () Is a function provided by the Keras API, which is a high-level deep learning library built on top of TensorFlow.Function takes in a list of sequences and performs the padding or truncation operation to make all sequences of equal length. It is commonly used in natural language processing (NLP) tasks.

Example:

sequence = [[1], [2, 3], [4, 5, 6]]

tf.keras.utils.pad\_sequences(sequence)

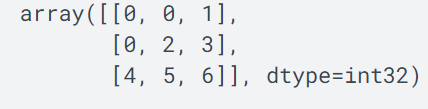




Fig 1.5

1. tf.keras.layers.Embedding () is a layer in TensorFlow's Keras API that is used for word embedding in natural language processing (NLP) tasks. It converts input integers (representing words or tokens) into dense vectors of fixed size. Can also be used in more complex network architectures, such as recurrent neural networks (RNNs) or convolutional neural networks (CNNs), where it serves as an input layer for the text data.
2. tf.keras.layers.GlobalAveragePooling1D () TensorFlow's Keras API that performs global average pooling over the temporal dimension of a 1D input tensor. It is commonly used in text classification or sequence-to-vector tasks where the input is a sequence of vectors, such as word embeddings or encoded sentences.GlobalAveragePooling1D layer is often used as a simple and efficient alternative to flattening the sequence and applying a Dense layer to process the entire sequence.
3. tf.keras.layers.Dense() s a fully connected layer in TensorFlow's Keras API. It is one of the fundamental building blocks of a neural network. The layer performs a linear transformation on its input.
4. tf.keras.models.Sequential () is a class in the TensorFlow library's high-level API, Keras. It represents a linear stack of layers, where you can add and configure layers in a sequential manner.

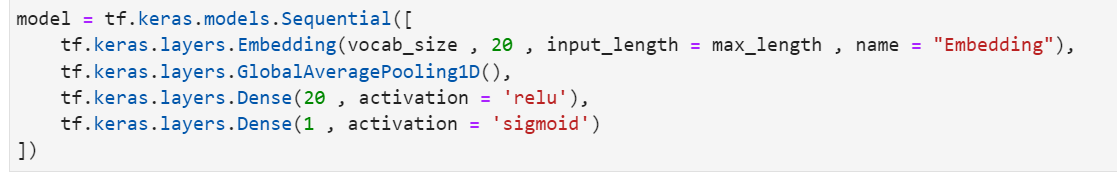


Fig 1.6

1. model.compile () is used to compile all the layers we fit into the model and by using certain parameters we calculate the loss and accuracy of the neural network. About **optimizer = ‘adam’**. Adam (short for Adaptive Moment Estimation) is a popular optimization algorithm commonly used in deep learning. It combines the advantages of two other optimization algorithms, AdaGrad and RMSProp, to provide efficient and adaptive optimization of the model's weights.



Fig 1.7

1. model.fit () is a method commonly used in machine learning frameworks, such as TensorFlow or Keras, to train a machine learning model on a given dataset.



Fig 1.8

Reference Link:

1. About Keras <https://keras.io/guides/sequential_model/>
2. Random Forest Model
3. Navie Bayes <https://www.javatpoint.com/machine-learning-naive-bayes-classifier>
4. Logistic Regression <https://www.w3schools.com/python/python_ml_logistic_regression.asp>
5. Gradient boosting <https://www.mygreatlearning.com/blog/gradient-boosting/>
6. Support Vector Machine <https://www.javatpoint.com/machine-learning-support-vector-machine-algorithm>
7. Sequential model by tensorflow <https://www.tensorflow.org/guide/keras/sequential_model>
8. Decision Tree Classifier <https://scikit-learn.org/stable/auto_examples/tree/plot_unveil_tree_structure.html#sphx-glr-auto-examples-tree-plot-unveil-tree-structure-py>