MINI PROJECT

(2020-21)

IMAGE CAPTION GENERATOR

Mid Term Report

Department of Computer Engineering & Applications

Institute of Engineering & Technology

Supervised by: **Submitted by:**
Mr. Mandeep Singh   Juhi Agarwal(181500301)
(Assistant Professor)   Parakh Ashra (181500449)
                                      Parth Maheshwari(181500456)
                                      Pankhuri Verma(181500448)

Acknowledgment

It gives us a great sense of pleasure to present the synopsis of the B.Tech Mini Project (IMAGE CAPTION GENERATOR) undertaken during B.Tech III rd Year. This project in itself is going to be an acknowledgement to the inspiration, drive and technical assistance will be contributed to it by many individuals.

We owe special debt of gratitude to **Mr. Mandeep Singh**, Assistant Professor Department of CEA, for providing us with an encouraging platform to develop this project, which thus helped us in shaping our abilities towards a constructive goal and for his constant support and guidance to our work. His sincerity, thoroughness and perseverance is been a constant source of inspiration for us. We believe that he will shower us with all his extensively experienced ideas and insightful comments at different stages of the project & also taught us about the latest industry- oriented technologies.

We also do not like to miss the opportunity to acknowledge the contribution of all faculty members of the department for their kind guidance and co-operation.
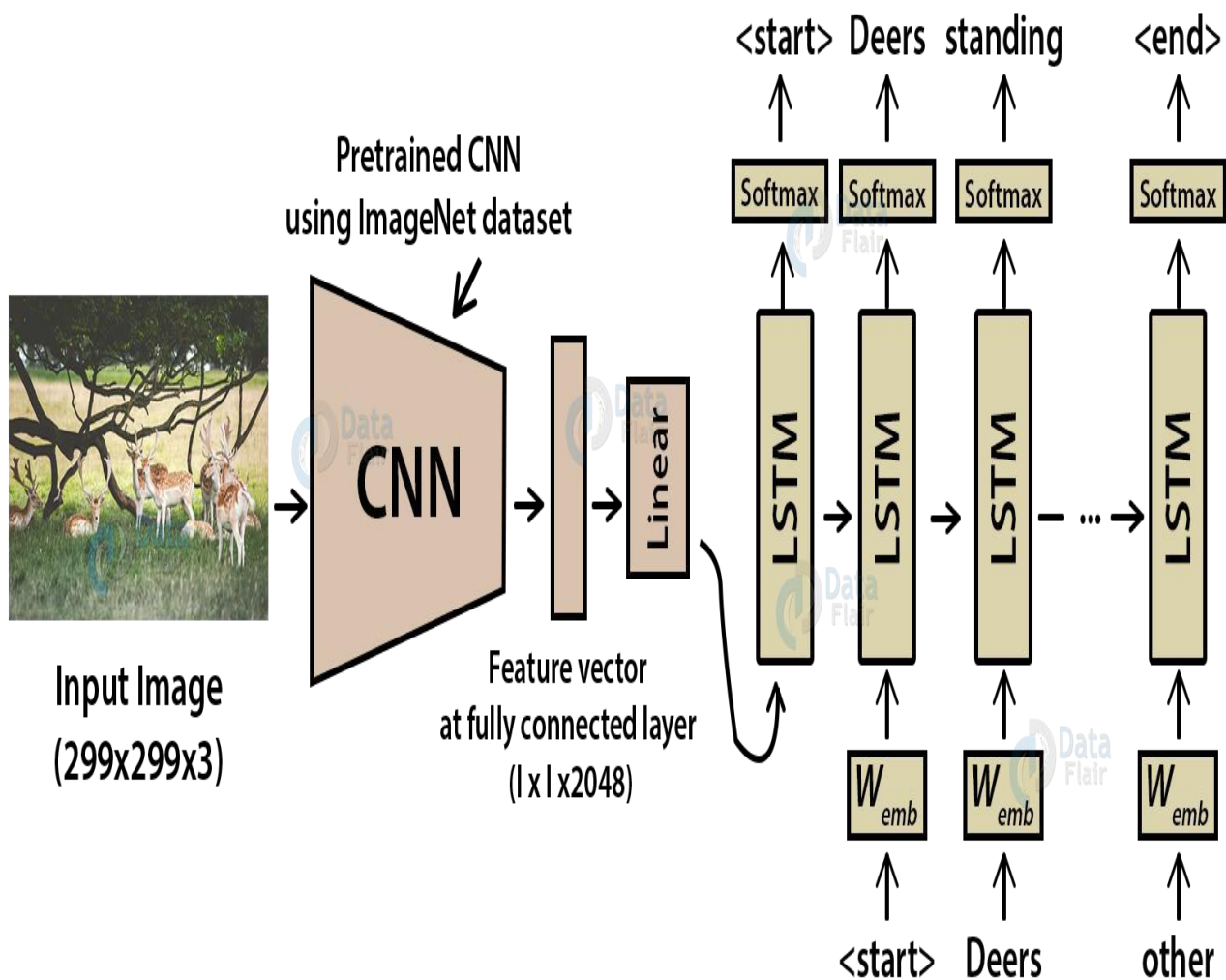
TABLE OF CONTENTS

Department of Computer Engineering & Applications, GLA University

IMAGE CAPTION GENERATOR:

Model - Image Caption Generator

INTRODUCTION

A. General Introduction of Topic

- What do you see in the below picture?

Well some of you might say "A white dog in a grassy area", some may say "White dog with brown spots" and yet some others might say "A dog on grass and some pink flowers"

The purpose of this project is to explain how Deep Learning can be used to solve this problem of generating a caption for a given image,hence the name Imagine Caption Generator!

B. Area of Computer Science :

- Computer vision researchers worked on this a lot and they considered it impossible until now! With the advancement in Deep learning

techniques, availability of huge data sets and computer power, we can build models that can generate captions for an image.

•We will use deep learning techniques of Convolutional Neural Networks and a type of Recurrent Neural Network (LSTM) together.

•We also use concepts of Natural Language Processing for textual representation of caption !

C. <u>Hardware and Software Requirements</u>

a) <u>Hardware (minimum)</u>:
  - External Hard Drives for Backup
    - Internet connection
    - 4GB RAM
    - Hard disk
    - i3 Processor (6th gen)
    - Minimum resolution :1024x765 Display

b) <u>Software</u>:
  - Tensorflow
  - Keras
  - Pillow
  - Numpy
  - Tqdm
  - jupyterlab

2. Problem Definition:

In recent years, with the rapid development of artificial intelligence, image caption has gradually attracted the attention of many researchers in the field of artificial intelligence and has become an interesting and arduous task. Image caption, automatically generating natural language descriptions according to the content observed in an image, is an important part of scene understanding, which combines the knowledge of computer vision and natural language processing.

The application of image caption is extensive and significant, for example, the realization of human-computer interaction. This project summarizes the related methods and focuses on the attention mechanism, which plays an important role in computer vision and is recently widely used in image caption generation tasks. Furthermore, the advantages and the shortcomings of these methods are discussed, providing the commonly used data sets and evaluation criteria in this field. Finally, this project highlights some open challenges in the image caption task.

3. __Objective:__

•A captioning model relies on two main components, a CNN and an RNN. Captioning aims at merging the two to combine their most powerful attributes i.e.

1. CNN's excel at preserving spatial information and images; and
2. RNN's work well with any kind of sequential data, such as generating a sequence of words. So by merging the two, you can get a model that can find patterns and images, and then use that information to help generate a description of those images.

4.   __Implementation Details__

The task of image captioning can be divided into two modules logically – one is an image based model – which extracts the features and nuances out of our image, and the other is a language based model – which translates the features and objects given by our image based model to a natural sentence.

For our image based model we usually rely on a Convolutional Neural Network model. And for our language based model we rely on a Recurrent Neural Network. We'll feed an image into a CNN.We will use the Flickr 8k dataset.This dataset contains 8000 images each with 5 captions.

So now CNN acts as a feature extractor that compresses the information in the original image into a smaller representation. Since it encodes the content of the image into a smaller feature vector hence,this CNN is often called the encoder.

When we process this feature vector and use it as an initial input to the following RNN, then it would be called decoder because RNN would decode the process feature vector and turn it into natural language.

5.   Future Prospects

As day by day, we are facing advancement in the field of technology, this Image caption generator can be a useful application to retrieve data of an image.Image captioning has various applications such as recommendations in editing applications, usage in virtual assistants, for image indexing, for visually impaired persons, for social media, and several other natural language processing applications.
As an example : Self driving cars — Automatic driving is one of the biggest challenges and if we can properly caption the scene around the car, it can give a boost to the self driving system.

i. Aid to the blind — We can create a product for the blind which will guide them travelling on the roads without the support of anyone else. We can do this by first converting the scene into text and then the text to voice. Both are now famous applications of Deep Learning. CCTV cameras are everywhere today, but along with viewing the world, if we can also generate relevant captions, then we can raise alarms as soon as there is some malicious activity going on somewhere. This could probably help reduce some crime and/or accidents.

iii.Automatic Captioning can help, make Google Image Search as good as Google Search, as then every image could be first converted into a caption and then search can be performed based on the caption.

Use Case Diagram
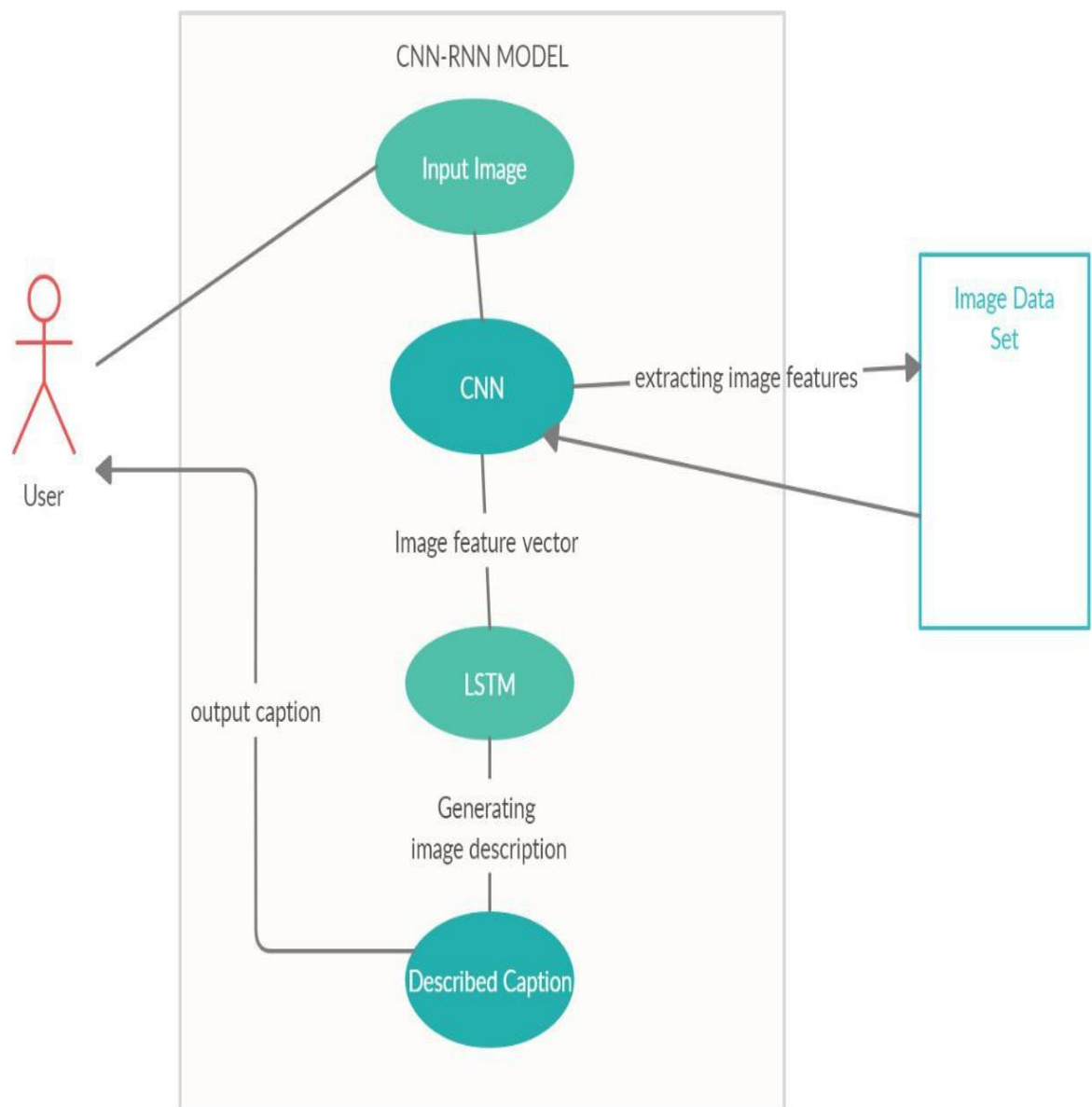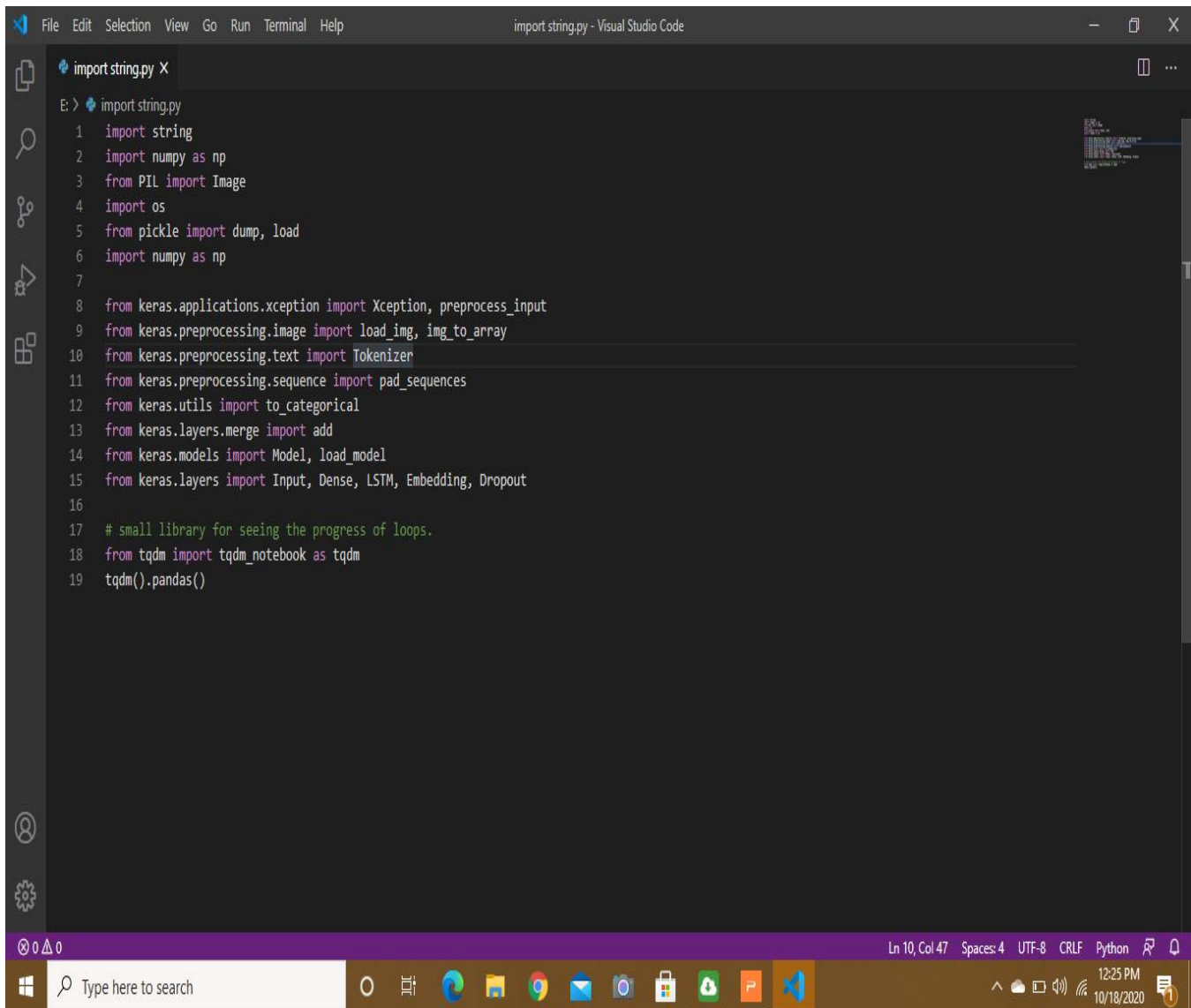
6.Progress till date:

1. We have successfully imported the packages and modules that will be used to make the Image caption generator.

2. Started the data cleaning process .

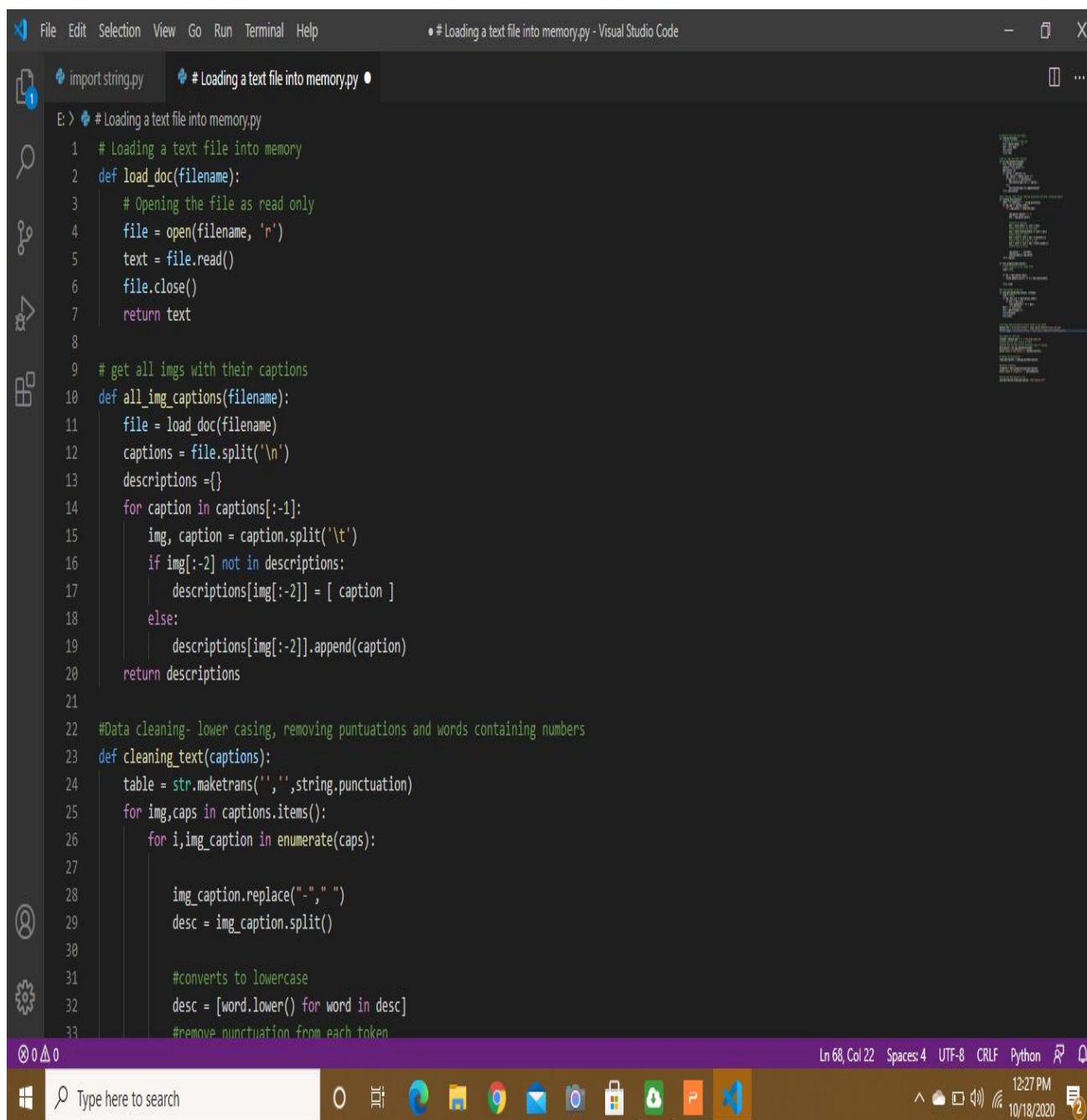3. We are learning the other concepts simultaneously as well.

Screenshots


```
File  Edit  Selection  View  Go  Run  Terminal  Help          import string.py - Visual Studio Code

import string.py ×

E: ›    import string.py
  1   import string
  2   import numpy as np
  3   from PIL import Image
  4   import os
  5   from pickle import dump, load
  6   import numpy as np
  7
  8   from keras.applications.xception import Xception, preprocess_input
  9   from keras.preprocessing.image import load_img, img_to_array
 10   from keras.preprocessing.text import Tokenizer
 11   from keras.preprocessing.sequence import pad_sequences
 12   from keras.utils import to_categorical
 13   from keras.layers.merge import add
 14   from keras.models import Model, load_model
 15   from keras.layers import Input, Dense, LSTM, Embedding, Dropout
 16
 17   # small library for seeing the progress of loops.
 18   from tqdm import tqdm_notebook as tqdm
 19   tqdm().pandas()
```

Fig 1 -Importing of modules and packages

Fig 2- Data cleaning and creation of descriptive text

Fig 3- Data cleaning and creation of descriptive text

File Edit Format Run Options Window Help

```
1000268201_693b08cb0e.jpg#0    A child in a pink dress is climbing up a set of stairs in an entry way .
1000268201_693b08cb0e.jpg#1    A girl going into a wooden building .
1000268201_693b08cb0e.jpg#2    A little girl climbing into a wooden playhouse .
1000268201_693b08cb0e.jpg#3    A little girl climbing the stairs to her playhouse .
1000268201_693b08cb0e.jpg#4    A little girl in a pink dress going into a wooden cabin .
1001773457_577c3a7d70.jpg#0    A black dog and a spotted dog are fighting
1001773457_577c3a7d70.jpg#1    A black dog and a tri-colored dog playing with each other on the road .
1001773457_577c3a7d70.jpg#2    A black dog and a white dog with brown spots are staring at each other in the
1001773457_577c3a7d70.jpg#3    Two dogs of different breeds looking at each other on the road .
1001773457_577c3a7d70.jpg#4    Two dogs on pavement moving toward each other .
1002674143_1b742ab4b8.jpg#0    A little girl covered in paint sits in front of a painted rainbow with her han
1002674143_1b742ab4b8.jpg#1    A little girl is sitting in front of a large painted rainbow .
1002674143_1b742ab4b8.jpg#2    A small girl in the grass plays with fingerpaints in front of a white canvas w
1002674143_1b742ab4b8.jpg#3    There is a girl with pigtails sitting in front of a rainbow painting .
1002674143_1b742ab4b8.jpg#4    Young girl with pigtails painting outside in the grass .
1003163366_44323f5815.jpg#0    A man lays on a bench while his dog sits by him .
1003163366_44323f5815.jpg#1    A man lays on the bench to which a white dog is also tied .
1003163366_44323f5815.jpg#2    a man sleeping on a bench outside with a white and black dog sitting next to h
1003163366_44323f5815.jpg#3    A shirtless man lies on a park bench with his dog .
1003163366_44323f5815.jpg#4    man laying on bench holding leash of dog sitting on ground
1007129816_e794419615.jpg#0    A man in an orange hat starring at something .
1007129816_e794419615.jpg#1    A man wears an orange hat and glasses .
1007129816_e794419615.jpg#2    A man with gauges and glasses is wearing a Blitz hat .
1007129816_e794419615.jpg#3    A man with glasses is wearing a beer can crocheted hat .
1007129816_e794419615.jpg#4    The man with pierced ears is wearing glasses and an orange hat .
1007320043_627395c3d8.jpg#0    A child playing on a rope net .
```

Fig 4- Data set

```
File  Edit  Format  Run  Options  Window  Help

1000268201_693b08cb0e.jpg        child in pink dress is climbing up set of stairs in
1000268201_693b08cb0e.jpg        girl going into wooden building
1000268201_693b08cb0e.jpg        little girl climbing into wooden playhouse
1000268201_693b08cb0e.jpg        little girl climbing the stairs to her playhouse
1000268201_693b08cb0e.jpg        little girl in pink dress going into wooden cabin
1001773457_577c3a7d70.jpg        black dog and spotted dog are fighting
1001773457_577c3a7d70.jpg        black dog and tricolored dog playing with each othe
1001773457_577c3a7d70.jpg        black dog and white dog with brown spots are starir
1001773457_577c3a7d70.jpg        two dogs of different breeds looking at each other
1001773457_577c3a7d70.jpg        two dogs on pavement moving toward each other
1002674143_1b742ab4b8.jpg        little girl covered in paint sits in front of paint
1002674143_1b742ab4b8.jpg        little girl is sitting in front of large painted ra
1002674143_1b742ab4b8.jpg        small girl in the grass plays with fingerpaints in
1002674143_1b742ab4b8.jpg        there is girl with pigtails sitting in front of rai
1002674143_1b742ab4b8.jpg        young girl with pigtails painting outside in the gr
1003163366_44323f5815.jpg        man lays on bench while his dog sits by him
```

Fig 5 - Descriptivetxt file.

● <u>REFERENCES</u>

◆ **Book References**
> ➢ Hands on ML with Scikit-Learn and Tensor-flow

◆ **Website References**
> ➢**https://towardsdatascience.com/image-_captioning-with-keras-teaching-computers-to-_describe-pictures-c88a46a311b8**
> ➢**https://data-flair.training/blogs/python-based-project-image-caption-generator-cnn/**

◆ **Faculty Guidelines**
> ➢ Mr. Mandeep Singh Sir