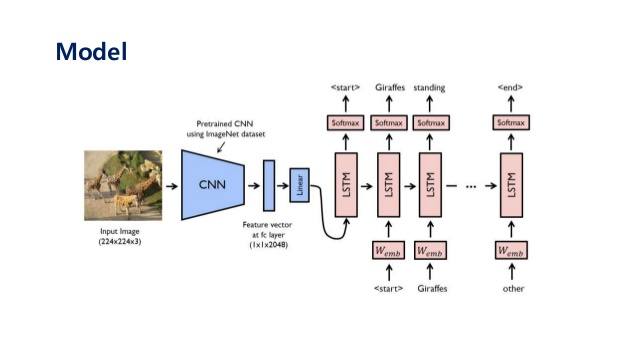
ARTIFICAL INTELLIGENCE PROJECT REPORT

# Image Caption Generator



**DONE BY:**

Arhan Ali Khan (11804301)

Abhay Pratap Singh Chauhan(11804223)

**SUBMITTED TO:**

Ankita Wadhawan ma’am

**Objective**: objective of this project is to achieve captions based on the dataset for the image given. now, we talk about what is image caption generator?Image caption generator is a task that involves computer vision and natural language processing concepts to recognize the context of an image and describe them in a natural language like English.

**METHOLOGY/ALGORITHM**

1. CNN

2. LSTM

**CNN:**

This section is done by Arhan Ali Khan(11804301)

Convolutional Neural networks are specialized deep neural networks which can process the data that has input shape like a 2D matrix. Images are easily represented as a 2D matrix and CNN is very useful in working with images.

CNN is basically used for image classifications and identifying if an image is a bird, a plane or Superman, etc.

**LSTM:**

This section is done by Abhay Pratap Singh Chauhan(11804223)

LSTM stands for **Long short term memory**, they are a type of RNN (**recurrent neural network**) which is well suited for sequence prediction problems. Based on the previous text, we can predict what the next word will be. It has proven itself effective from the traditional RNN by overcoming the limitations of RNN which had short term memory. LSTM can carry out relevant information throughout the processing of inputs and with a forget gate, it discards non-relevant information.

### The Dataset

For the image caption generator, we will be using the Flickr\_8K dataset. There are also other big datasets like Flickr\_30K and MSCOCO dataset but it can take weeks just to train the network so we will be using a small Flickr8k dataset. The advantage of a huge dataset is that we can build better models.

* [Flicker8k\_Dataset](https://github.com/jbrownlee/Datasets/releases/download/Flickr8k/Flickr8k_Dataset.zip)
* [Flickr\_8k\_text](https://github.com/jbrownlee/Datasets/releases/download/Flickr8k/Flickr8k_text.zip)

The Flickr\_8k\_text folder contains file Flickr8k.token which is the main file of our dataset that contains image name and their respective captions separated by newline(“\n”).

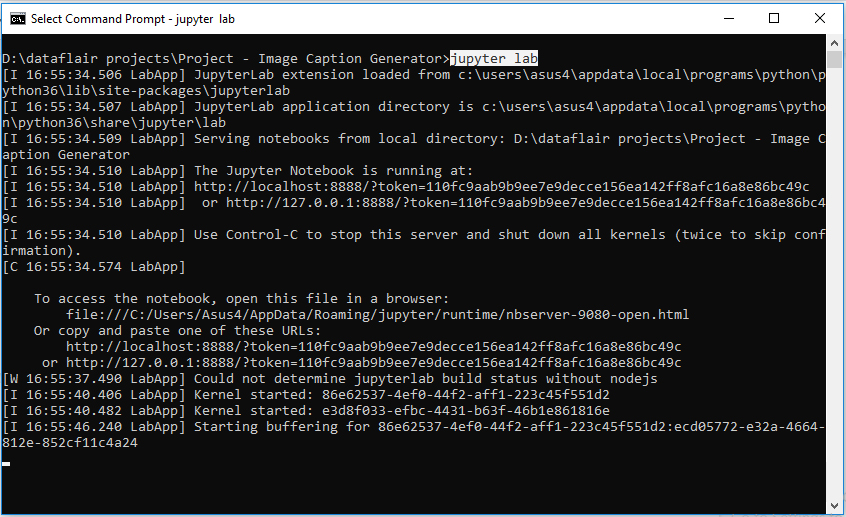
### Pre-requisites

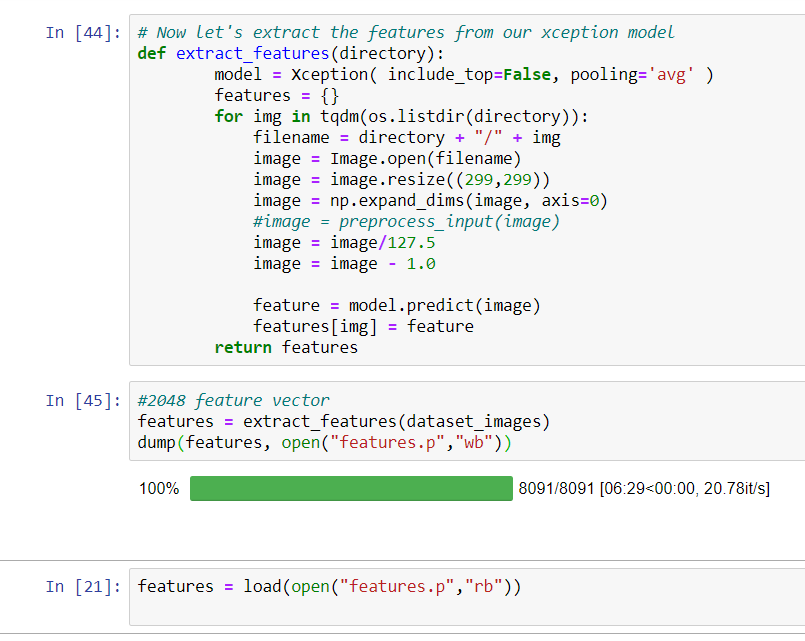
This project requires good knowledge of Deep learning, Python, working on Jupyter notebooks, Keras library, Numpy, and [***Natural language processing***](https://data-flair.training/blogs/nlp-natural-language-processing/).

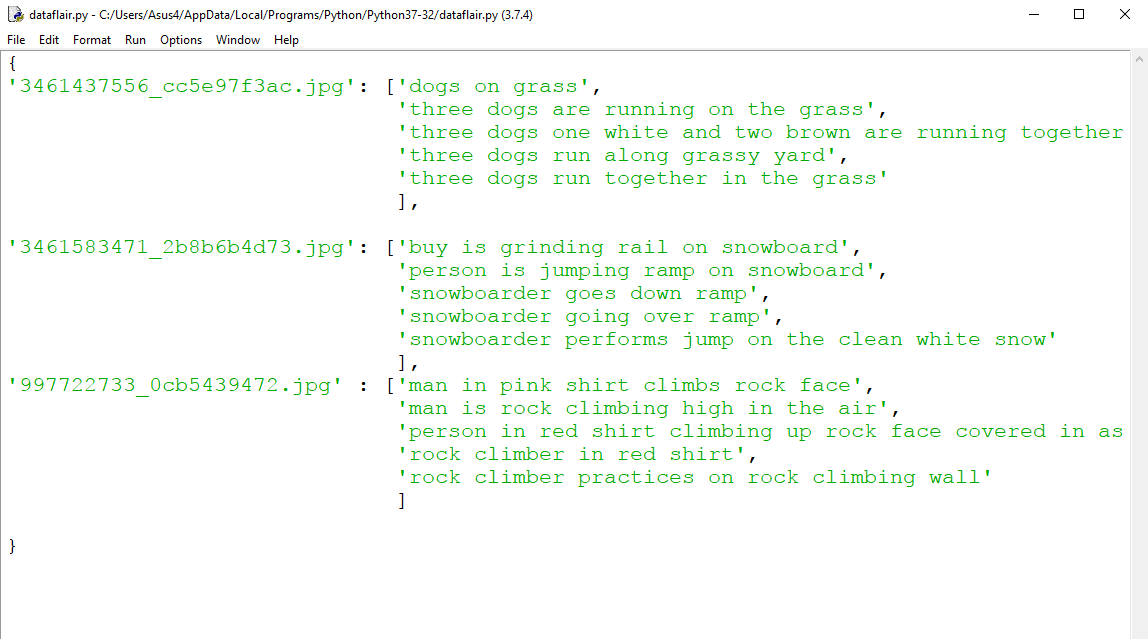
we have installed all these libraries:

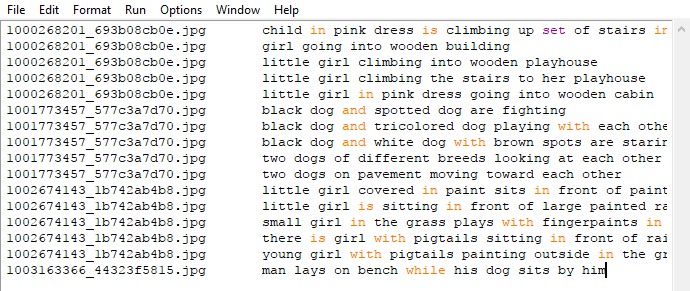
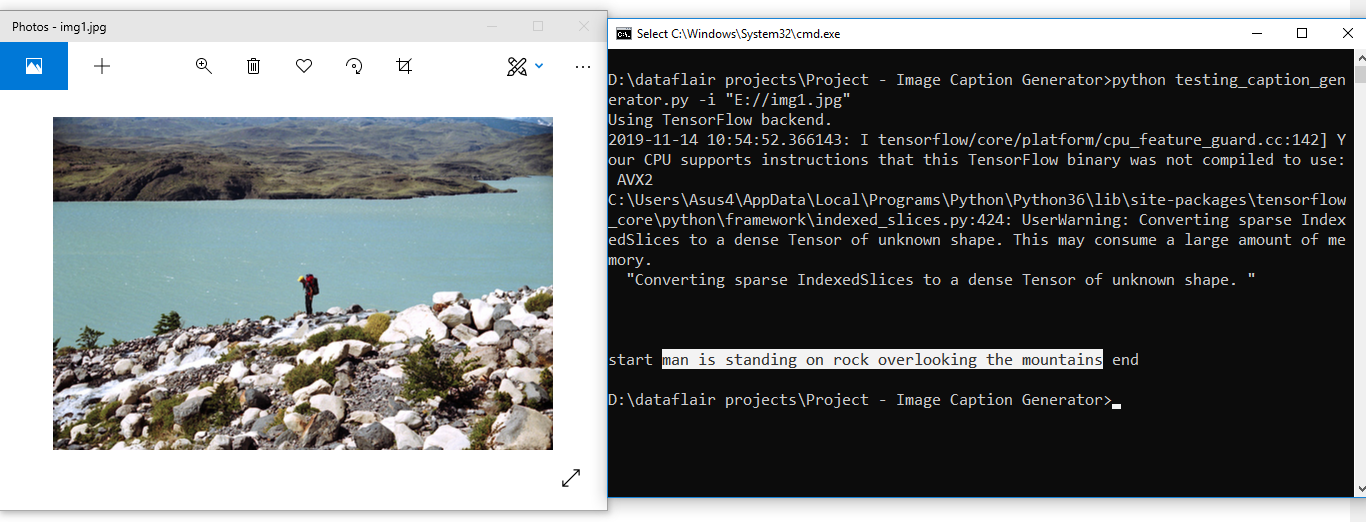
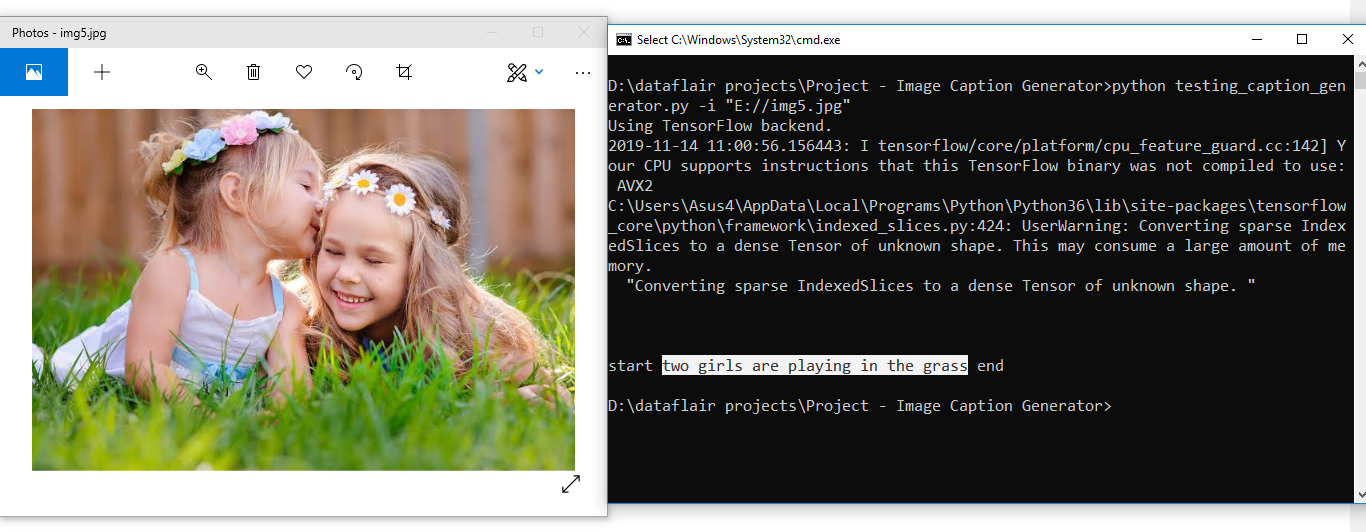
* pip install tensorflow
* keras
* pillow
* numpy
* tqdm
* jupyterlab

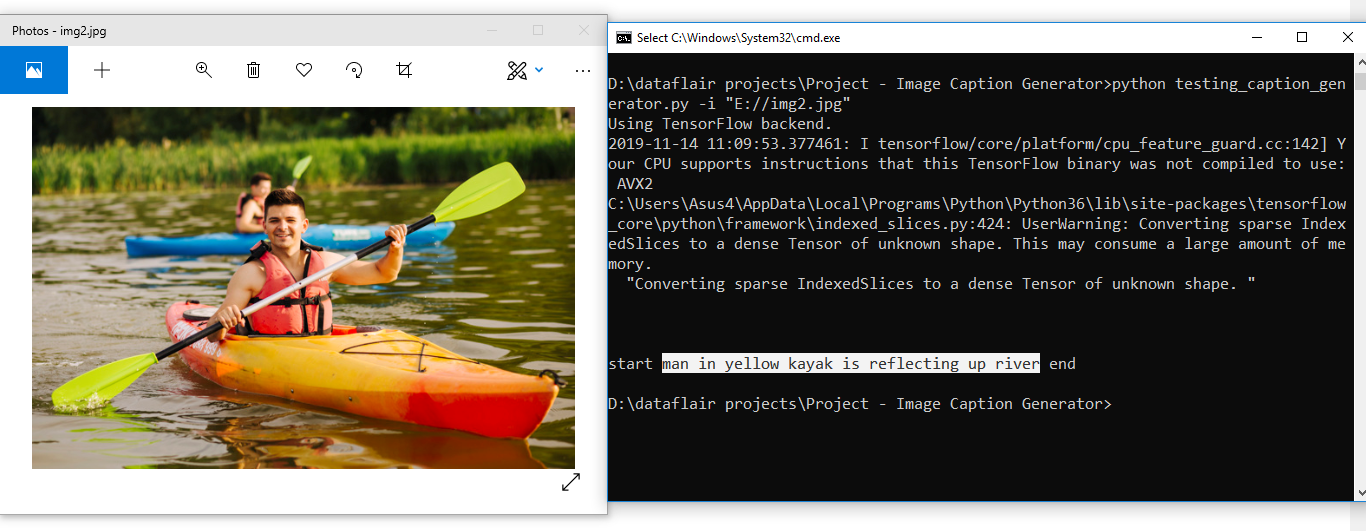
SCREENSHOTS OF OUTPUT and Procedures:











## Summary

In this Python project, we have implemented a CNN-RNN model by building an image caption generator. Some key points to note are that our model depends on the data, so, it cannot predict the words that are out of its vocabulary. We used a small dataset consisting of 8000 images. For production-level models, we need to train on datasets larger than 100,000 images which can produce better accuracy models.

THANK YOU