# Project Proposal

## Data Collection

**(a) Provide details on the data sources you will be using for your project. Are you using public datasets, collecting your own data, or obtaining data from specific sources?**

We are using the "Flicker8k_Dataset" and "Flickr_8k_text" for our project. These are open source datasets from Kaggle, and are used in the field of computer vision and natural language processing for tasks related to image description.

- Flicker8k_Dataset is a collection of 8,000 images. Each image is paired with five different captions, providing descriptions of the entities and events depicted in the image.

- Flickr_8k_text is related to the descriptions or annotations for the images in the Flicker8k_Dataset. It contains the textual descriptions that are associated with the images, providing the necessary context.

## Data Preprocessing

**(b) Preprocess your data using Data Cleaning, Transformation, and Feature Selection techniques.**

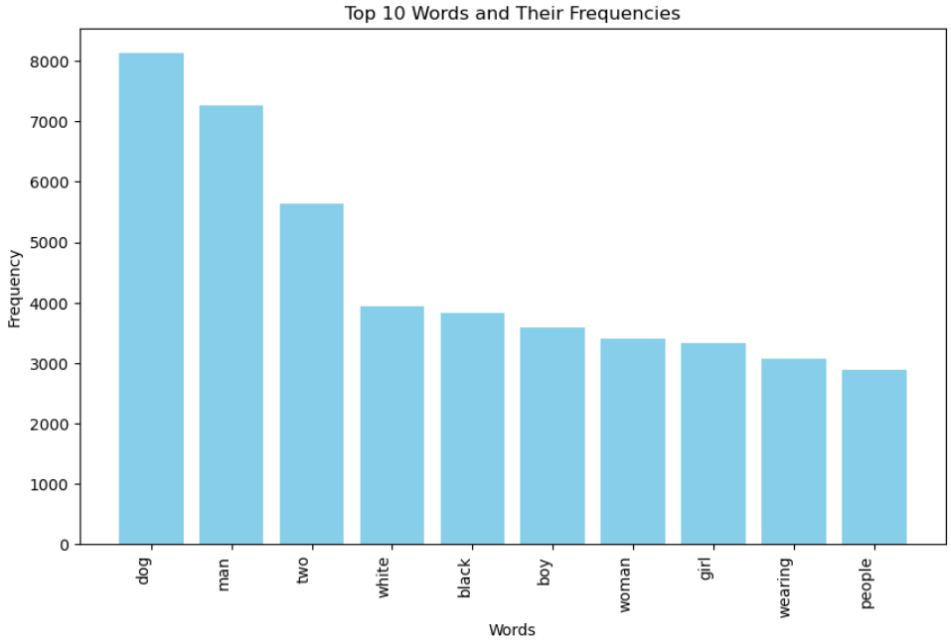
- We perform data cleaning by converting text to lowercase, removing punctuation, handling special characters, and eliminating non-alphabetic words.
- We perform transformation by dividing the data into tokens form with respect to each image. In addition, we split the captions into individual words and build a vocabulary of all the unique words in the captions.
- We perform feature extraction by using the Xception model, capturing vital visual information from the images. This extraction aids in feature selection, focusing on the most relevant aspects for the model to learn during training.

# Data Analysis and Visualization

**(d) Create visualizations to gain insights into the data, identify patterns, and help with feature selection and model understanding.**

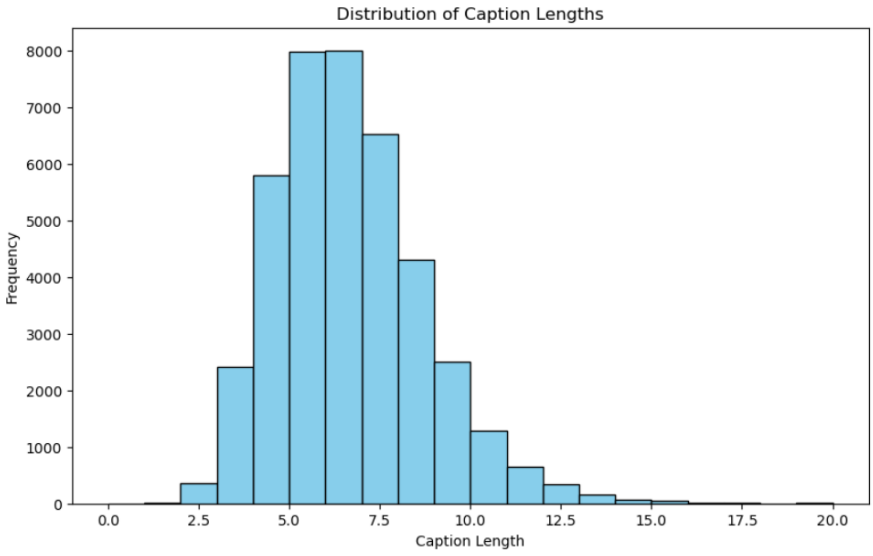
We get some insights of data by plotting data with different parameters using Matplotlib.

- Word frequency distribution: This visualization shows the most common words in the captions.




- Caption length distribution: This visualization shows the distribution of caption lengths.

```
Average Caption Length: 6.18 words
Maximum Caption Length: 20 words
Minimum Caption Length: 0 words
```



- Image-caption pairs: This visualization shows examples of image-caption pairs from the dataset.

"man in black shirt is playing guitar."

"construction worker in orange safety vest is working on road."

"two young girls are playing with lego toy."

## Model Selection

**(e) Research and select suitable AI models/algorithms for your project. Identify technical requirements for the project.**

We aim to use CNN for image feature extraction and an LSTM network for generating captions.

- Convolutional Neural Networks (CNNs): We will use CNNs for image feature extraction, capturing spatial hierarchies within images. Moreover, we aim to use pre-trained models like Xception, VGG, or ResNet for this purpose. These models are available in libraries such as Keras and TensorFlow.
- Long Short-Term Memory networks (LSTMs): We will use LSTM to generate sequential text (captions) from the extracted image features. We will use Keras library as it offers LSTM layers and functionalities for sequence generation.

For the successful execution of the project, we need to have a high level proficiency in Python language to execute and implement machine learning models. A profound understanding of deep learning concepts and practices is essential. This knowledge is particularly crucial when working with neural networks and large-scale models, which are fundamental components of our project's machine learning aspects. Moreover, we also need to be familiar with the Natural Language Processing (NLP) concepts. Lastly, familiarity with the libraries such as TensorFlow, Numpy, Pillow and Jupyter NoteBooks etc is essential for training and executing our models.