**NLP Case Study**

**Image Captioning in Telugu and Tamil**

**Team Members:**

|  |  |
| --- | --- |
| **Roll Number** | **Name** |
| CB.EN.U4CSE19302 | Akash Guna |
| CB.EN.U4CSE19311 | Bharath Anuj P |
| CB.EN.U4CSE19322 | G Surya Santosh Reddy |
| CB.EN.U4CSE19329 | K Siddhartha Reddy |

**Abstract:**

Image Captioning is a task where we write a short description provided an image. When an image with multiple objects is shown to humans each would have a different perspective on the image hence results in different captions. Caption generation from images could be used to gain knowledge on a Deep Learning model’s perspective on the image. We build an Attention Based Image Captioning System which would preserve the model’s perspective. Captions being language specific can deter the performance of an Image Caption Generator. We therefore compare the usage of Tamil and Telugu caption for accurate Image Caption Generators. We Use Flickr 8k Dataset and translate its captions to Tamil and Telugu using Google Translation services. The Image Captioning Model used in this project is a combination of CNN and LSTM networks. CNN is used as a Feature Extractor for extracting the semantic and object information from the input image and to produce attention maps. These Information are passed into a LSTM for predicting the Image Captions

**Dataset generation:**

**Step – 1:**

Downloaded Flickr8 dataset from Kaggle. This dataset contains images and its captions in English.

**Step -2:**

Translate the English captions into Telugu and Tamil using google translator API.

**Sample image from dataset:**



**English caption given:**

A child in a pink dress is climbing up a set of stairs in an entry way

**Telugu caption generated:**

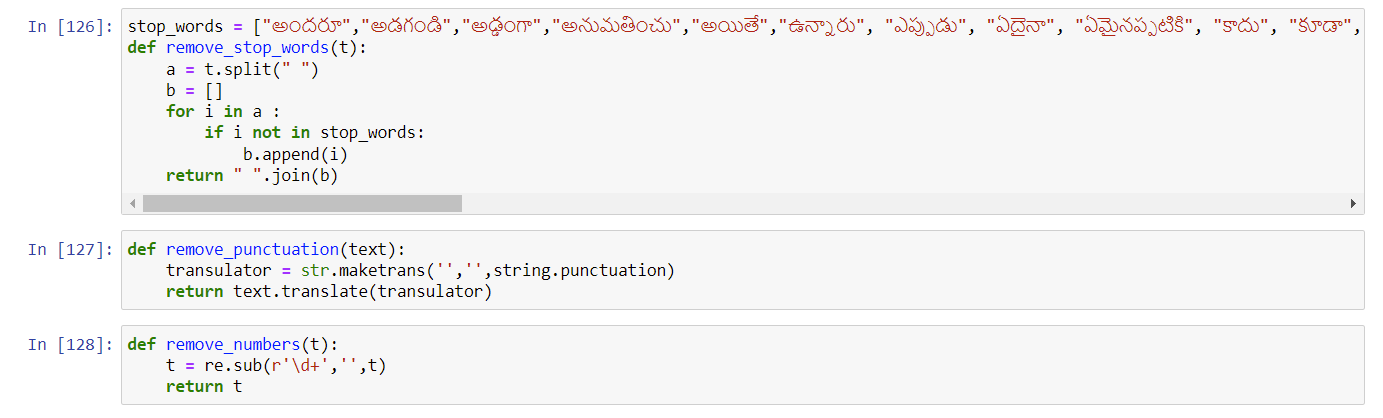
పింక్ డ్రెస్‌లో ఉన్న పిల్లవాడు ప్రవేశ మార్గంలో మెట్ల సెట్ పైకి ఎక్కుతున్నాడు.

**Tamil caption generated:**

இளஞ்சிவப்பு நிற உடையில் ஒரு குழந்தை நுழையும் வழியில் படிக்கட்டுகளில் ஏறுகிறது.

**Preprocessing used:**

* **Removing punctuation**: Since the punctuations do not contribute anything to the data while applying the NLP techniques, we remove them from the data.
* **Removing numbers:** Numbers won’t give much importance to get the key terms or entities.
* **Removing stop words:** The stop words do not hold any significant meaning in the data. So, removing them would make the entity recognition easier.

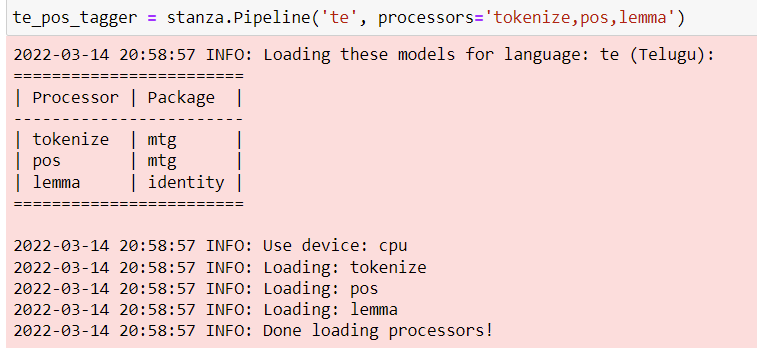


After preprocessing, tokenization, lemmatization, POS tagging is done using Stanza API

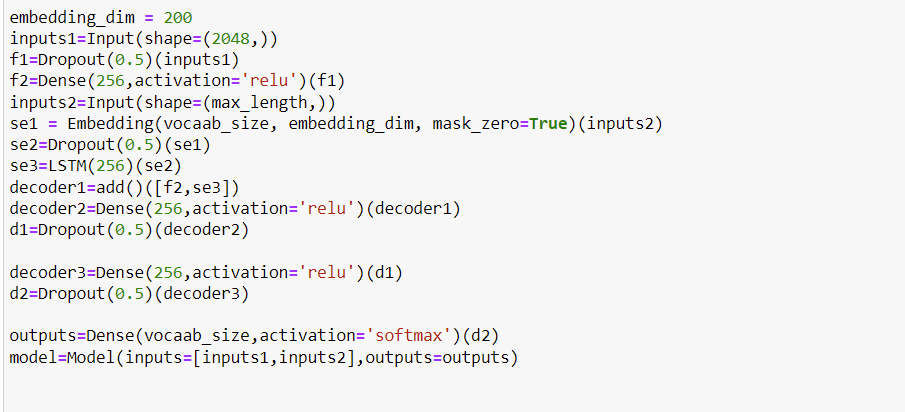
* **Lemmatization:** It is a method which reduces the words to root form like stemming but this approach considers the context of the word and gives a meaningful base form.

### Parts Of Speech Tagging:

* **Keyword ‘tag’:** Tagging is a kind of classification that may be defined as the automatic assignment of description to the tokens. Here, the descriptor is known as tag, which represents one of the part-of-speech, semantic information.
* **Definition:** The Process of tagging/assigning/categorizing words in a text(corpus) with its corresponding Parts of Speech, depending on the definition of the word and its context based on its occurrence is called

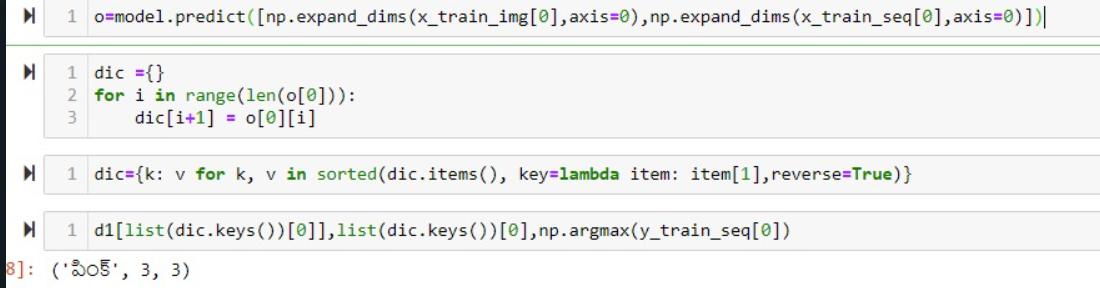


**Architecture:**



**Results**

**Telugu output:**



**Tamil output:**

