**Image Caption Generator**

OBJECTIVE: To predict the captions for the input image

DOMAIN: Computer Vision , NL , LLM

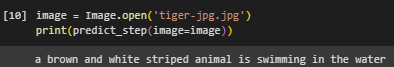
DURATION: 1 month

PLAN OF ACTION:

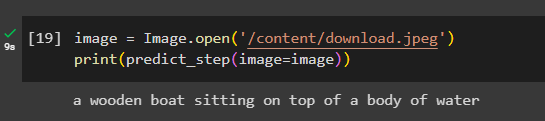
* ~~Data Collection and Preprocessing~~
* ~~Model training~~
* ~~Testing and Validation~~

If possible after the development of the basic model we plan to extend this project for the benefit of blind individuals, the generated image captions could be transformed into auditory descriptions through text-to-speech technology, providing an accessible means for them to comprehend and engage with visual content.

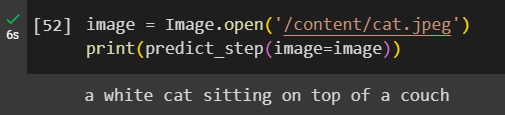




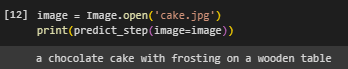




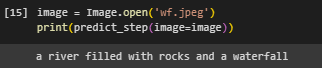




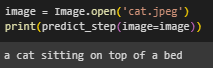




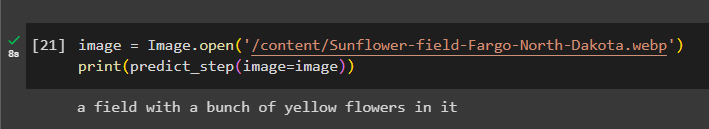












Dataset - Flikker 8K

Model - Hugging Face

Reference

<https://youtu.be/aaP7JJZuvGs?si=Mdwf2yMBgdi_eDi->

<https://youtu.be/8PzcmL9d3zM?si=ba-iV4Q_JzF7o0zS>

**2nd plan of action**

* Improve the model (Pytorch)
* Use more dataset like MS COCO Dataset
* Text to speech

**UPDATE**

* Added the Text to speech method using Google Translate API