

# CS263: NLP Final Project Proposal

## Team Members

- Simran Masand 706085004
- Shruthi Srinarasi 705900515
- Vasu Eranki 706072935
- Chinmay Gurjarpadhye 505852105

## Flickr Image Dataset: Image to Prompts

Our task for this challenge is to predict the captions that were annotated on the images. We expect to leverage existing visual transformer models and fine-tune models used for image captioning. We will choose one of the more popular benchmarks e.g. Nocaps and COCO; BLEU or CIDER metric. We have access to 30,000 images which are 256x256.

<https://www.kaggle.com/datasets/hsankesara/flickr-image-dataset>

## List of Deliverables:

- Literature Survey of existing models
- Comparative study for baseline models and state of the art models.
- Training and Testing Results
- Evaluation and Conclusion

## Division of Work

- Choosing and evaluating relevant dataset - SM/SS
- Image Processing - SS/VE
- Literature Review - VE/CG
- Implementation of various models - CG/VE
- Evaluation of implemented models - SM/VE
- Reporting and analysis - VE/VE