

## PROJECT 1

**Title:** Automatic Image Captioning

**Objective:** Build an image captioning model to generate captions of an image using CNN

**Dataset Link:** [Flickr8k\\_dataset](#)

**Dataset description:** A collection of sentence-based image description

- Dataset consists of 8k images in JPEG format with different shapes and sizes.
- Images are paired with five different captions which provide clear descriptions of the salient entities and events.
- The images were chosen from six different Flickr groups and included a variety of scenes and situations.

**Project Overview:** Captioning the images with proper description is a popular research area of Artificial Intelligence. A good description of an image is often said as “Visualizing a picture in the mind”. The generation of descriptions from the image is a challenging task that can help and have a great impact in various applications such as usage in virtual assistants, image indexing, a recommendation in editing applications, helping visually impaired persons, and several other natural language processing applications. In this project, we need to create a multimodal neural network that involves the concept of Computer Vision and Natural Language Process in recognizing the context of images and describing them in natural languages (English, etc). Deploy the model and evaluate the model on 10 different real-time images.

**Tools:** Natural Language Toolkit, TensorFlow, PyTorch, Keras

**Deployments:** FastAPI, Cloud Application Platform | Heroku, Streamlit, Cloud Computing, Hosting Services, and APIs | Google Cloud

**Final Submissions:**

- Project Technical Report
- Project Presentation with desired outcomes
- Summary of 3 research papers

## PROJECT 2

**Project\_2A Title:** Classifying multi-page documents

**Objective:** Classify text with one or more classes or categories of a document for making it easier to manage and sort.

**Dataset Link:** [BBC\\_News](#)

**Dataset description:** BBC news website collection of stories from five topical areas

- Consists of 2225 documents from the BBC news website for the years 2004 and 2005.
- Class Labels: 5 (business, entertainment, politics, sport, tech).

**Project Overview:** Text is a rich source of information, but extracting insights from text is challenging and time-consuming due to the unstructured data. With the growing scale of text data in industrial applications, automatic text classification is becoming increasingly important. Also separating information from the irrelevant can reduce the cost and time of searching and retrieving information from all the documents. In this project, we are aiming to build a model which can categorize the documents to make them easier to manage and sort.

**Tools:** Natural Language Toolkit, TensorFlow, PyTorch, Keras

**Deployments:** FastAPI, Cloud Application Platform | Heroku, Streamlit, Cloud Computing, Hosting Services, and APIs | Google Cloud

**Final Submissions:**

- Project Technical Report
- Project Presentation with desired outcomes
- Summary of 3 research papers

## PROJECT 2

**Project\_2B Title:** Classifying multi-page documents

**Objective:** Classify email messages into user-specific folders based on the fields such as “From”, “Subject”, “Body”, and “To, CC”.

**Dataset Link:** [Enron\\_Email\\_Dataset](#) (May 7, 2015 Version of dataset)

### **Dataset description:**

- The Enron email dataset contains approximately 500,000 emails generated by employees of the Enron Corporation.
- It was obtained by the Federal Energy Regulatory Commission during its investigation of Enron's collapse.
- This is the May 7, 2015 Version of the dataset, as published at <https://www.cs.cmu.edu/~enron/>

**Project Overview:** Automated classification of email messages into user-specific folders and information extraction from chronologically ordered email streams have become interesting areas in text learning research. In this project, we focus on the problem of assigning messages to a user’s folders based on the fields such as “From”, “Subject”, “Body”, and “To, CC”.

**Tools:** Natural Language Toolkit, TensorFlow, PyTorch, Keras

**Deployments:** FastAPI, Cloud Application Platform | Heroku, Streamlit, Cloud Computing, Hosting Services, and APIs | Google Cloud

### **Final Submissions:**

- Project Technical Report
- Project Presentation with desired outcomes
- Summary of 3 research papers

## PROJECT 3

**Title:** Image tagging and road object detection

**Objective:** Detect object tagging in the video and examine how parallel object detection on multiple patches can allow the detection of smaller objects in the overall image without decreasing the resolution.

**Dataset Link:** [BDD 100K Dataset](#).

**Dataset description:** The Berkeley Deep Drive (BDD) dataset is one of the largest and most diverse video datasets for autonomous vehicles.

- The dataset contains 100,000 video clips collected from more than 50,000 rides covering New York, San Francisco Bay Area, and other regions.
- The dataset contains diverse scene types such as city streets, residential areas, and highways.
- Furthermore, the videos were recorded in diverse weather conditions at different times of the day.

**Project Overview:** Object detection and segmentation methods are one of the most challenging problems in computer vision which aim to identify all target objects and determine the categories and position information. Numerous approaches have been proposed to solve this problem, mainly inspired by methods of computer vision and deep learning. In this project, we aim to build a model which detects multiple objects and segmentation in a moving video. For eg. Image tagging, lane detection, drivable area segmentation, road object detection, semantic segmentation, instance segmentation, multi-object detection tracking, multi-object segmentation tracking, domain adaptation, and imitation learning.

**Tools:** TensorFlow, PyTorch, Keras

**Deployments:** FastAPI, Cloud Application Platform | Heroku, Streamlit, Cloud Computing, Hosting Services, and APIs | Google Cloud

**Final Submissions:**

- Project Technical Report
- Project Presentation with desired outcomes
- Summary of 3 research papers

## PROJECT 4

**Title :** Automatic Speech Recognition(ASR)

**Objective:** Build an ASR model for converting speech to text.

**Dataset Link :** [LibriSpeech](#)

**Dataset description:** LibriSpeech is a corpus of reading English speech, suitable for training and evaluating speech recognition systems, published in 2015 by Johns Hopkins University. It is derived from audiobooks that are part of the LibriVox project and contains 1000 hours of speech sampled at 16 kHz of 2000 speakers. The LibriVox project<sup>1</sup>, a volunteer effort, is responsible for the creation of approximately 8000 public domain audiobooks, the majority of which are in English. Most of the recordings are based on texts from Project Gutenberg<sup>2</sup>, also in the public domain. The data is already divided into train/dev/test sets. The total size of the data is 60 GB and subsets are available of different sizes.

Initially, we recommend working only with 'dev-clean' and 'test-clean' datasets for building the model. We can use any one or a combination of both data sets as a training set. A subset of either 'dev-clean' or 'test-clean' can be used for testing purposes. Once modeling is done with these smaller data sets, start modeling using 'train-clean'/'train-other' data sets of larger sizes as a training set. Now, 'dev-clean', 'test-clean', and 'test-other' datasets are used for validation/testing purposes only.

**Project Overview:** Automatic speech recognition is the application of Machine learning or AI where human speech is processed and converted into readable text. We can find numerous applications such as Instagram for real-time captions, Spotify for podcast transcriptions, youtube video transcription, Zoom meeting transcriptions, etc. The field has grown exponentially over the last few years. An explosion of applications taking advantage of ASR technology in their products to make audio and video data more accessible.

There are different approaches to Automatic Speech Recognition, viz. traditional HMM (Hidden Markov Models) and GMM (Gaussian Mixture Models) and end-to-end deep learning models. In this project, we aim to build and deploy a model that can generate the written text from the speech with a decent accuracy.

**Tools:** TensorFlow, PyTorch, Keras, Audio Processing tool/library.

**Deployments:** FastAPI, Cloud Application Platform | Heroku, Streamlit, Cloud Computing, Hosting Services, and APIs | Google Cloud

**Reference:** [Papers using libriSpeech](#)

**Final Submissions:**

- Project Technical Report
- Project Presentation with desired outcomes
- Summary of 3 research papers