**Projects**

**Summarization of Online Meeting Conversations using NLP & Attention mechanism :** (*August 2023 – present)*

* + This is an end to end project where we convert the meeting video/audio into text using a speech to text converter.
  + For this purpose, we will make use of “**Whisper**” pre-trained model by **OpenAI** to generate the text from speech.
  + The text thus generated will then be summarized using **seq2seq models** which make use of encoders, decoders, **BERT** and

**attention mechanism**. We will be using 2 pre-trained models to perform these 2 tasks.

* + Both these pre-trained models will be fine-tuned based on our dataset and additionally, we are looking at training the

models in a QA manner (Question and Answer based training) which helps the model to learn faster.

**Bangalore House Price Prediction using Linear Regression :** *(May 2022 – Aug 2022)*

* + It is an end to end House Price Prediction Project, we have used machine learning techniques to predict the price of houses.
  + This project involves analyzing historical data to predict the sale price of a house in Bangalore city using Linear Regression

based on various features such as location, size, number of bedrooms, and other relevant factors.

* + Feature selection and engineering performed to improve the accuracy of the model.
  + Selected suitable features for building a regression model and our trained model will predict the prices on our test data.
  + The accuracy of the model will be validated by comparing predicted prices with actual prices, and once deemed accurate,

the model will be deployed in a web application to predict house prices based on their features.

[Github\_Project\_](https://github.com/anupsubbu98/House_Price_Prediction_ML)

**Spam Classification using NLP :** *(Aug 2022 – Sept 2022)*

* + Developed a machine learning model for spam classification project using NLP techniques.
  + Preprocessed the data by removing stop words, stemming the words, and converting the text into numerical features.
  + Trained the model on a dataset of labeled emails using *Bag of Words* and *TF-IDF* techniques.
  + Machine learning model was trained using various algorithms, including *Logistic Regression, SVM, Random Forest* etc.
  + Evaluated the model using various performance metrics such as accuracy, precision, confusion matrix.
  + Classified new emails as either spam or ham based on their content using the developed model.

[Github\_Project](https://github.com/anupsubbu98/Spam-Classifier)

**Cricket Data Analytics & Exploratory Data Analysis :** *(Feb 2022 – May 2022)*

* + Cricket Data Analytics project is made on IPL and International criket data till 2019.
  + We can analyse the data of the matches and get some statistical insights to predict the performance of the players on

that particular match.

* + Performed Exploratory Data Analysis (EDA) to gain insights into patterns, relationships, and trends in data for

data science projects.

* + Leveraged EDA insights for developing predictive models or making informed decisions in business and research.
  + Utilized EDA techniques to provide valuable insights for advancing the field of data science.

[Github Project](https://github.com/anupsubbu98/Kaggle-Competitions)

*Key results :*

* Participated in Data Preprocessing Techniques in order to make data useful for creating Machine Learning models.
* Created pipelines which includes Data Collection, Feature Engineering, Feature Selection, Model Creation,

Hyperparameter Tuning and Model Deployment.

* Built various regression and classification algorithms by using various Sklearn libraries such as Linear Regression,

Logistic Regression, Decision Trees, SVM, Ensemble techniques.