FINAL REPORT

PROJECT: GRAMMAR AUTO-CORRECT

DOMAIN: MACHINE LEARNING

PROJECT's INTENT : Building a model to take English sentences as input from the user and analyse them and identify the grammatical error's and rectify them and return sentence with correct grammar and also suggest's a impressive way of writing those sentences which have a better impact on reader.

INITIAL APPROACH: Firstly, we researched various hugging face transformers and read various research papers on grammar auto-correct to generate precise corrected sentence. In our process we tried various hugging face models. We aimed to understand the necessary tech stacks and devised a tentative project timeline:

WEEK 1: Read various research paper and discuss about their pros and cons.

WEEK 2: Fine-tune and train transformer models for Grammar auto-correct.

WEEK 3: Modify models for Paraphrase text, Improved grammar text, Simplified text, Cohesive text, and Neutralized text and taking audio input.

WEEK 4: Finalize the project, creating a Gradio interface.

FINAL ITERATION AND/OR METRICs: We used rouge score as our metric to calculate the efficiency of our model to correct the text.

TIMELINE:

Week 1-2:

Under the guidance of our mentors, we learnt NumPy, PyTorch, Pandas, MatplotLib, Scikit-Learn, Tensorflow, Hugging Face NLP course.

Week 3-4:

DATASET FINDING AND EXPLORATION

We started our hunt for various datasets and pre trained models that are smaller to gain hands on experience. After checking Metrics like Rouge Score , analysing the quality of results. We also built our own custom built dataset and used other datasets like coedit small and large.

We tried many different models and T5 was best, but due to limitations of computer resources we used t5 small model for fine-tuning on the datasets.

We fine-tuned T5 small model on our custom built dataset and coedit.

Week 5-6:

WORKING ON MODIFICATIONS

We added various modifications like Paraphrase text, Improved grammar text, Simplified text, Cohesive text, and Neutralized text and taking audio input.

Week 7:

Finally, we combined all using pipeline and successfully designed our website using Gradio.

here is our final code:
https://colab.research.google.com/github/varshneyanushka/IITI_SOC_ML9/blob/main/Mahabharat.ipynb

SCOPE OF IMPROVEMNTS-

1. We tried adding audio input but because of time limitation it couldn't be completed.
2. We can also add input through camera and convert it into text.