Sumit Tagadiya

+91 7043105881 | tagadiyasumit@gmail.com | in sumit-tagadiya | ⌂ sumittagadiya | sumittagadiya.medium.com

EDUCATION

K J Institute of Engineering and Technology (GTU)   Vadodara, Gujarat
Bachelors in Computer Engineering (CGPA - 8.60 Out of 10)   *Aug 2016 - May 2020*

TECHNICAL SKILLS

Main Skills: Data Analysis, Machine Learning, Deep Learning, NLP, Computer Vision
Languages: Python, SQL (Postgres), JavaScript, HTML
Frameworks: Tensorflow, keras, Scikit-learn, Flask
Developer Tools: Linux, Git, Docker, Google Cloud Platform, VS Code, Heroku, AWS
Libraries: Pandas, NumPy, Matplotlib, SciPy, NLTK,BeautifulSoup

PROJECTS

Stack Overflow Search Engine based on Semantic meaning   Sept 2020 – Oct 2020
- Developed a Search Engine on Stackoverflow Data based on the semantic meaning of the question. When a user searches any question then the search engine would give the most similar results based on semantic meaning rather than keyword matching.
- The main task of this project is to understand the content of what the user is trying to search for and then return the most similar results in minimal time based on the user's query using semantic similarity.
- Glove pre-trained vectors, TF-IDF, Gensim Word2Vec model, and Tensorflow Hub module have been used for sentence encoding. To similarity measure, Cosine similarity has used.
- Deployed the whole project on Heroku and made it end to end for a better experience.
- Published a Blog on medium explaining my approach to this project

English to Italian Machine Translation with Vanilla and Attention Mechanism   Oct 2020 – Nov 2020
- Developed a model using Vanilla Encoder-Decoder and Attention mechanism with LSTM
- Implemented Research paper of Bahdanau's Attention Mechanism with three scoring functions named Dot Scoring, General Scoring, and Concat Scoring from Scratch
- To compare the results of all these Scoring functions and vanilla Encoder-Decoder BLUE score has been used for performance measure.

Image Segmentation on Indian Traffic Data   Nov 2020 - Dec 2020
- Implemented CANET model by referring Attention-guided Chained Context Aggregation for Semantic Segmentation research paper from scratch.
- For Performance measure, IOU(Intersection Over Union) Score has been used.Model will identify 21 different objects from image/video.

Document(Text) Classification using CNN   Aug 2020 - Sept 2020
- Trained two Deep CNN models one on word level embedding and second on character level embedding.
- In this task Data set contains 20 News groups text data so model predicts text into 20 different classes, Also Too much Preprocessing was required in this Task.
- Pre-trained Glove Vectors with 300d has been used for text embeddings.Model built using keras functional API.

CNN on CIFAR Dataset   July 2020 - Aug 2020
- Implemented Densenet Model by referring Densely Connected Convolutional Networks research paper   and trained it from scratch.
- Densenet architecture is made by 2 blocks named Dense block and Transition block. Output block contains Batch Normalization, GlobalAvgPooling followed by Softmax layer.
- By implementing Densenet Architecture i got 90% test accuracy which is pretty high.

Microsoft Malware Detection   June 2020 - July 2020
- Implemented Multiclass classification Model using XGBoost Algorithm to classify different malware files into respected classes. There were total 9 Nine malware classes.
- Dataset Size was too large, Almost around 200 GB. Out of which 50GB of data was bytes files and 150GB of data was asm files.
- The main objective of this project is to minimize multiclass log loss as much as possible. I got a minimum log loss of 0.001. To get this much minimum log loss I made some high-level features like image feature from byte file and asm file etc. I have also used some feature selection techniques like the Chi-Square test etc.