



DOPPELGÄNGER

SYNTHETIC DATA GENERATOR

INDIVIDUAL REPORT FOR **ANKEIT TAKSH (A0213496B)**

GRADUATE CERTIFICATE IN PATTERN RECOGNITION SYSTEMS – PRS GROUP 19

The project was a brainchild of ours with special credit to Nirav for the new idea ,after multiple discussion and considering the real world requirement of data scientists to search for data. Hence we collaborated to ensure that the a solution can be achieved wherein dataset can be produced to serve as fuel for the data science industry.

Research work: Our research work was via towardsdatascience , medium articles , official tensorflow, and github projects websites .Though the project as a whole cannot be broken down to subsections which elaborates the specific portions for individuals but on my part I was more focussed on the solution implementation ,deployment and testing part with the Algorithms which again was Continuous integration and continuous deployment mode. Aside from production grade product implementation , generating of image synthetic data using CNN and analysis of the outcome was the most important part where we tested various ML techniques . We had used the GANs, VAE , Autoencoders. I with team had worked on many existing codes mostly from Kaggle and github and how we can utilize the existing modules to leverage on our project.

Development: Worked with team with focus on GAN and Autoencoder part using python tensorflow to ensure Synthetic image generation with high accuracy and same can be validated. We tested on tensorflow and realized the quality of data output for GAN was better than VAE. So we brought in a combo of GAN and VAE. Worked with our hardcore ML guy(Anirban) to test the results wherein he can focus on coding part. As well worked with our project planner and UI developer (Prashant) to host everything over the public domain so the solution is globally available.

Architecture: Actively involved with team in ensuring that final product had all the actual production level design was deployed. We dockerized the whole solution to implement machine learning ops. This enabled us to achieve the level of enterprise grade solution as companies are struggling to implement machine learning right now.

Test Cases: Worked with team for test cases scenarios with different datasets like mnist and celeba to cross check the results across various data format.Applied the test case scenario with different methods definitely with the awesome team and had a great learning for same.

Project documentation:

Investigated the National data protection papers of EU and documented the same . Actively translated the results to project doc and added the graphical comparative view. Cross Validation analysis and model construction part has been done and helped team to consolidate all in project documentation.



DOPPELGÄNGER

SYNTHETIC DATA GENERATOR

Analysis and Learning:

The project application thus enabled us to explore various methods of not just encoding and generation with one algorithm but to cross verify results across another algorithms. This can be used implemented in various sections of machine learning.

--End--