Akhil Vydyula

Email: akhilvydyula1111@gmail.com **Mobile number :** +91-8074191867

PROFESSIONAL SUMMARY

- Self motivated and smart working Data Science engineer, passionate about cutting edge technologies and solving real world problems . Eager to convert data to business achievements.

- I am a result focused Data Scientist with experience in developing robust code for real world problems. Interested in devising solutions for challenging tasks, learning and applying new technologies and tools. An enthusiastic team player with a can-do attitude and a strong user focus.

SKILLS AND COMPETENCIES

- **Operating Systems :** Windows, Linux
- **Programming Languages :** Python, SQL, HTML, CSS, JavaScript
- **Areas of Interest :** Machine Learning, Deep Learning, NLP, Computer Vision, Data Analysis.
- **Tools/Libraries :** Tensorflow, Keras, Pandas, Numpy, Seaborn, Matplotlib, OpenCv
- **Database :** MySQL, SQLite3
- **Frameworks and Deployments :** Flask, Heroku
- **Version control :** Git

BUSINESS EXPERIENCE

Trainee Machine Learning Engineer – Atos Syntel,india

Oct 2019 – Present

- Fabricated training dataset for deep learning models using Node.js and Python which improved the performance by 50-60 percent.
- Created a Developer Portal to facilitate customers with product usage.
- Trained a custom named entity recognizer (NER) which enabled detection of named entities and helps in extracting the necessary information more precisely.
- Worked on servers and Docker in the production environment.independently worked on developing the front end of the client's website using HTML, CSS,javascript.

PROJECTS

Quora Insincere questions classification(Machine learning,Deep learning, Sklearn, Pandas, Numpy, Seaborn, Matplotlib)

- Developed a model to predict whether a question asked on Quora is sincere or not.
- Built Attention layer and lstm models,CNN, Logistic regression baseline model.
- Using CNN model received loss: 0.1005 on train data and val_loss: 0.104, Applying Attention model and lstm were able to improve test loss to 0.090.For Logistic regression accuracy is 0.506, Random forest model is 0.50, LSTM is 0.688, Attention layer 0.672,CNN 0.6686.
- The model can be trained for Questions in Quora,It will automatically say Insincere or not.

Santander Customer Transaction Prediction (Machine learning, Sklearn, Pandas, Numpy, Seaborn, Matplotlib)

- Santander is one of the leading Commercial Banks based in Madrid, Spain. The problem statement was to predict whether a customer will do transactions in future or not.
- Those predictions were done using the previous transaction details.
- There were 200 anonymised features in the dataset, I had done the complete project starting from the EDA , followed by feature engineering till training of various machine learning models and selected the best model out of them as the final model.
- The final model had the AUC score of 90.041% using the encoder part and it produces a context vector. The encoded sequence is decoded with the help of context vectors in the decoder part.

Question Pair Similarity (Machine Learning, Seaborn, Python, Pandas, Numpy, Matplotlib, NLP)

- This project was done using the Kaggle's Quora question pair dataset.
- The aim was to identify which questions asked on Quora are duplicates of questions that have already been asked. This could be useful to instantly provide answers to questions that have already been answered.
- We were tasked to predict whether a pair of questions were duplicate or not. The performance is evaluated on log-loss value. I performed EDA, text cleaning, feature engineering, text preprocessing and model building.
- I started with a baseline model which was a simple random model that classified data randomly into similar or non-similar targets. Later the more sophisticated models' performance was judged using this baseline model. Xgboost gave the best log-loss value of 0.364 on the test data.

StackOverflow Tag Prediction (Machine Learning, Seaborn, Matplotlib, Python, Pandas, Numpy, NLP)

- StackOverflow is the largest, most trusted online community for developers to learn, share their programming knowledge, and build their careers.
- This problem statement was asked in the Facebook Recruitment Challenge on Kaggle.
- The business objective was to predict as many tags as possible with high precision and recall. There were no strict latency constraints. It was a multi-label classification problem.
- The performance metric was Micro-Averaged F1-Score. EDA , followed by text cleaning, preprocessing and training of various models was done.

Personalised Cancer Diagnosis (Machine learning, Seaborn, Matplotlib, Python, Pandas, Numpy)

- This problem statement was brought by Memorial Sloan Kettering Cancer Center (MSKCC) to take the personalised medicine to its full potential.
- In the problem, we had to classify the given genetic variations/mutations based on evidence from text-based clinical literature into 9 different classes.
- The interpretability of the decisions were important in problems related to medical datasets. Thus, it was one of the constraints in this case also. We were provided with two data files one of which contained the information about the genetic mutations and the other contained the clinical evidence (text) that human experts/pathologists use to classify the genetic mutations. Both these data files had a common column called ID.
- The performance metrics were multi-class log-loss and confusion matrix. To get the proper understanding of data, EDA was performed which was later followed by text cleaning to remove

Virtual Drawing Pad using OpenCv (OpenCv, Python, Numpy)

- This project was about tracking the path of an object. Here I used OpenCv to track a blue coloured object using colour as the key component. It can be used as an interface to various machine learning and deep learning applications.

SOCIAL PROFILES

Github | LinkedIn | Portfolio