Prakhar Gandhi
Github | Linkedin | my site | Portfolio | gprakhar0@gmail.com | +917339830157

Education

July 2015 - Dec 2019 B.Tech. (Bachelor's of Electronics and Instrumentation) from **Birla Institute of Technology and Science, Pilani, Hyderabad Campus, India** (GPA: 7.15/10.0)

Coursework : Data Structures and Algorithms, Neural Networks and Fuzzy Logic, Discrete Math for CS, Digital Image Processing
Self Taught Courses: Database Management Systems, , Number Theory: Project Euler
Office Courses : Deep Learning: Advanced NLP and CNNs, RNNs, LSTMs and GRUs, Bayesian Machine Learning in Python: A/B Testing, Recommender Systems and Deep Learning in Python, AI and Meta-Heuristics (Combinatorial Optimization) Python, Certificate Link
Udemy : Deep Learning Prerequisites: Linear Logistic Regression in Python, Data Science: Supervised Machine Learning in Python, Cluster Analysis and Unsupervised Machine Learning in Python, Data Science: Deep Learning and Neural Networks in Python, Modern Deep Learning in Python, Machine Learning and AI: Support Vector Machines in Python, Ensemble Machine Learning in Python: Random Forest, AdaBoost, Data Science: Natural Language Processing(NLP) in Python, Machine Learning: Natural Language Processing in Python(V2), Natural Language Processing with Deep Learning in Python, Unsupervised Deep Learning in Python, Deep Learning: Advanced Computer Vision(GANs, SSD + More!) Relevant BITS Course Link

Technical Skills

- **Languages** Python, C++, JavaScript, SQL, CSS, HTML, Latex, Object Oriented Programming
- **Databases** PostgreSQL, MongoDB, Cloud Firestore, Firebase
- **Machine Learning** Scikit-learn, Pandas, Numpy, Keras, Tensorflow, Pytorch, SciPy, NLTK, matplotlib, seaborn, Probability and Statistics
- **Tools and Frameworks** Jupyter Notebook, Flask, Django, pySpark, Git, Jira, Confluence, Dash, Bitbucket

Experience

Standard Chartered Bank GBS - Data Analyst,Associate Projects (Data Science) Aug 2020 - present

- **Python Scripting (Banking) | Python, pandas, numpy, pdoc, opencv**
  - Helped automate manual data extraction processes by doing **Keyword extraction of e product documents**(client ids, and other confidential details) using **python and regex**. This work improved SLA for the agreed deliverables from **1 day** to **1 minute** per document.
  - Automated synthetic dataset generation by doing **Multi blending labels of product documents**(watermarks, barcodes, seals and other confidential details) using **python and opencv**. This work involved Parallel Processing improved Data Creation for the agreed deliverables from **4 hour** to **1 hour** per document 256 combinations of logos.
  - Detected **mathematical noise patterns** and implemented the functions for same using **SOLID Principles using OOPS**.
  - Other works involve NER Explainability Pipeline creation and image duplicate prediction.
  - Learnt how to efficiently query the data from a database using PostgreSQL.
- **Django REST API Creation | Python, postgresql, pandas, numpy**
  - Created apis that generates datapoints from 3 main ML Pipelines(ML,CV and NLP) and keep it into postgresql database..
  - Created tables in database, optimized code to put million rows of csv files into tables.
  - Made various apis for text features generation like adverb count, noun count, adjective count, profanity checks , etc from csv, excel, text and pdf files.
  - Made various apis for image features generation like average brightness score, mean brightness score, average blurrness score, skewness check, etc from single and multiple images .
  - Made functions for Upper Control Limit and Lower Control Limit using Inter Quartile Range for discrete data and X-bar and s control charts values for continuous data. Send these calculations to frontend to show in chart under Model Monitoring and Improvement Section.
  - Daterange and classwise statistical accuracy and stability charts datapoints are sent in UI to visualize the drift happening overall.
  - Also created an api for payslips image to text conversion and generated both image and text features from the text.
  - Testing:- Used query params from postman app as input for testing purpose.
  - Deployment:- Deployed all of the APIs in the linux virtual machine.
  - Architecture:- implemented a api response that includes pagination loop to show limited amount of data in one component and uses the concept of the same to get the next batch of data using the same api response but change in queries.
  - Communicated and presented insights clearly and compellingly to senior leadership of the organization

Projects

- **Text Robustness app | Python, Flask, Spacy** Office private project
  - Developed a flask app to visualize model predictions and flipped classes based on synonym replacement, insertion and deletion of tense words of banking words present in a general spacy dictionary to check the robustness of model to any such attacks.
- **NER/POS Tagging Web Application | Python, Spacy, numpy, pandas, Flask, Dash, HTML/CSS, JS** Office Private Project
  - Used Spacy models to tag all words of paragraph of a text column in csv datasets to their Named Entity or Part of Speech.
  - Automated an web interface using Flask API to interact with the model and finally deployed app.
- **Bayesian Optimization to increase the metrics of bad performant model | Python,Hyperopt,LazyPredict** Link to Demo
  - Developed best performant globally optimized models from the list of models whose accuracy is far lower from baseline set accuracy(can include regression case as well and other metrics like f1 score can be taken as well).
- **Noise Pattern Detector Deployed App | Python, Opencv, Dash** Link to Demo
  - Implemented web app for noise detection in images using continuous wavelet transform with total noisy pixels count, deployed at render and visualized the various types of noises and finally reduced the uploading time by making it work only for resized version of original image.
  - Same idea can be scaled to detect other unknown types of noises by using reconstruction error and different normalized distance metrics like l2 norm.
- **Avengers Face Detector Deployed App | Python, Sklearn, dlib, face-recognition** Link to Demo
  - Implemented avengers face detection app in Python using machine learning with encoded feature vector generation of training dataset, deployed at huggingface and reduced the response time to 10 ms and finally achieved 95% accuracy.
  - Same idea can be scaled to detect a thousands of faces by optimizing the code using numpy vectorization techniques.

Achievements

- max 3* rating on codechef, Link.
- max 1368 rating on codeforces, Link.
- created a 3 layer artificial neural network from scratch as a part of research paper provided to me in coding assignment from the knowledge gained from udemy courses and gained 96% accuracy for benchmarks like iris dataset.(code can be run for mnist dataset as well), Link.
- converted Matlab to Python code for all of the ML algorithms from scratch as a part of BITS Course NNFL in various coding assignments from the knowledge gained from udemy courses and gained 81%-96% accuracy for assigned datasets.(included stacked autoencoders with backprop), Link.
- helped a senior colleague in ML Robustness Visualizations, Poison Detection using ML Techniques