Prakhar Gandhi
Github | Linkedin | my site | Portfolio | gprakhar0@gmail.com | +9133234325

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 , 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**.
  - Implemented **multiprocessing** on django microservices by using parallel processing using **pandarallel** package and decreased the execution time.
  - Implemented **multiprocessing** at hackathon on Large Language models using transformers to generate pairs of questions and answers on a set of company pdf files.
- **Ner Explainability Creation | Python, sklearn, pandas, numpy, plotly**
  - Researched on methods to detect explainability of an existing system at word level accuracy of 80%.
  - Implemented NLP pipeline for a ML based NER explainer search engine. Worked on text preprocessing functions and integrated them in pipeline to prepare input for a machine learning model during training as well as prediction.
  - Deployed local level explainability on trained SKLEARN CRFSUITE model using dash app on localhost interface to handling and improving query classwise visualization response for a single sentence accurately.
  - Created Champion Challenger and added NER pipeline for existing developed pipelines by comparing champion log files and challenger log files for all of the best selected metrics for ML with regression and classification, Object detection with/without classification and Text Prediction models using python. Reduced overall time complexity to consider Million points(scalable). Further selected best metrics for Object detection pipelines.
  - Communicated and presented insights clearly and compellingly to senior leadership of the organization
- **Image duplicate prediction (Banking) | Python, pandas, numpy, tensorflow(imagededup), pytorch(detecto), opencv**
  - Developed analytical and ML models to assist other insider client team in identifying the right duplicates and influential missing documents for various banking pdfs to improve the overall duplicate detection of barcodes, seals and other important contents of the TradeLC Docs by atleast 80%.
  - Built ML models to get probability of bounding box detection of barcodes, watermark seals, handwritten signatures (Used Structural Similarity Index (SSIM) score on top of cosine similarity score before comparison of two images) Projected the bounding box to dataset and retransformed the data to clients. Achieved 85% Accuracy score using transfer learning of faster rcnn mobilenet models on neural nets.
  - Worked on data visualization, report creation of analytical models to support our hypothesis and for client presentation.
  - Selected best metrics for object detection with or without classification and deployed same into production in champion challenger pipeline.

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

- red coder on atcoder. Link
- 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