Prakhar Gandhi

micro1 certified

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

Gold Award Have received a gold award from Standard Chartered Bank Global Business Services Private Limited for a project done as a part of Hackathon

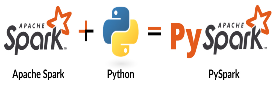
Technical Skills

- **Languages** Python, C++, JavaScript, SQL, CSS, HTML, Latex, Object Oriented Programming, Data Structures And Algorithms, Data analysis
- **Databases** PostgreSQL, MongoDB, Cloud Firestore, Firebase
- **Machine Learning** Scikit-learn, Pandas, Numpy, Keras, Tensorflow, Pytorch, SciPy, NLTK, matplotlib, seaborn, Probability and Statistics, decision trees, survival analysis, time series analysis, Neural networks
- **Tools && Frameworks** Jupyter Notebook, Flask, Django, pySpark, github, Jira, Confluence, Dash, Bitbucket , Azure functions

Experience

Standard Chartered Bank GBS - Senior Data Analyst (Data Analyst II),Associate Projects (Data Science) Aug 2020 - Jan 2024

- **Python Scripting (Banking) | Python, pandas, numpy, pdoc, opencv**
  - managed 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(t5-large model) to generate pairs of questions and answers on a set of company pdf files.
  - Implemented **McNemar Test via statistical analysis** for cross-team members from structured data on ML Pipeline to compare the performance of 2 classifiers and selecting the best model further integrating it within the champion challenger pipeline.
- **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. Developed 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 (implemented 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.
  - Developed 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.



Experience with Pyspark

- **Pyspark Zycus assignment**
  - Built an automated text classification pipeline to classify products on the basis of text reviews provided by the users. Github
  - Improved data processing speed by 91% by creating a parallel computing pipeline to convert a column of paragraphs into [tokenizer to stopwords remover to vectorizer to inverse document frequency to logistic regression model] and get the final predictions.
  - Filtered out the labels in code to remove **NaN labels** to improve the final evaluation metrics and finally applied the performance metrics like accuracy, weighted precision, weighted recall, area under curve and f1 score to obtain the final verdict.

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

○ Implemented 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.

● Image Global Explainability | Python,Networkx,Dash Youtube Link to Demo

○ Developed an dash app for a similar class visualization of a particular image in networkx graph based on custom features to visualize the connections.

● Automated Topic Modelling App using BERT | Python,BERTopic,Dash Youtube Link to Demo

○ Developed an dash app for a topic modelling visualization of the text column in pandas dataframe in the context of unsupervised learning(can include supervised case as well).

● 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.
8 Kyu
prakhr
AtCoder
[Change Photo]
Country/Region India
Birth Year 1996
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.

● created chatbots in python that answers the query of integral calculus and differential calculus in form of equations, Link.

● helped a senior colleague in visualizing robustness check of ML/DL model at a targetted classwise image level attack and made a opacity test to compare a permutation of two classes at a time by implementing the same from the paper. Further Detected poison in image dataset. Paper Link.

Small Startup project work July 2018 - Dec 2018

● Backend Development/Python Development/Software Development/Web Scraping/Documentation generation

○ Built an automated monitoring system to transfer one project to another over time on NoSQL cloud firestore db.Github
○ Managed**webscraping** to scrape kirana datasets of sites like justdial.com over multiple metric definitions and org-ids to increase scalability of our own product by 150 times.Extracted the hidden elements of phone numbers of dynamically loading scripted websites like justdial with lxml parsing.
○ Automated **sessions, cookies, and web authentication** mechanisms via python code while logging into mobile friendly version of facebook website.
○ Performed and Simulated user interactions on the web pages of facebook.
○ Performed Google search Automation to find missing product info.
○ Navigated through Firewalls, User Logins, and requirements of facebook mbasic website using random time sleep function to fool the website into thinking there is no web-crawler crawling the website.
○ Developed code to extract all the friends of user on facebook and learnt the use of **recursion** to support daily runs on production of barcodes by reducing the time complexity using memoization technique and created **tkinter responsive app** for the same that uses the concept of **multithreading**.

Big Achievements as Hobby(TOP PROJECTS)

● HANDWRITTEN CHARACTER RECOGNITION Link ● ANOMALY DETECTION Link ● AUDIO TO WORD CLOUD Link
● AUTOMATE EXCEL CREATION Link ● AUTOMATE SQL QUERY Link ● CALCULATE NET MONEY Link
● CREATE CLASSIFIER PIPELINE Link ● MULTIVARIATE ANALYSIS HYPOTHESIS TESTING Link
● VISUALIZE DATASET(ONE-LINER CODE) Link ● DETECT DEADLOCK Link ● DRIFT CALCULATOR Link
● DRONE SIMULATOR Link ● AUTOMATE IMPUTATION Link
● GARBAGE TEXT CLEANER Link ● IMAGE TO TEXT PIPELINE Link ● OPERATE SPATIAL NETWORK Link
● FAIRNESS PIPELINE Link
● RECONSTRUCT IMAGE Link ● RECONSTRUCT TEXT Link ● REMOVE NOISE Link
● SIMPLEST AUTOMATED CLI REGEX CHECKER Link ● TIMESERIES DATA METRICS Link ● TRANSLATE TO ALL Link
● TRANSLATE ENGLISH FROM CHINESE Link ● TRANSLATE SENTENCES Link ● TRAIN CHATBOT(MATHBOT) Link
● OUTLIER DETECTION TEXT PIPELINE Link ● OUTLIER DETECTION TABULAR PIPELINE Link
● OUTLIER DETECTION IMAGE PIPELINE Link ● OUTLIER DETECTION AUDIO PIPELINE Link
● SCALABLE TEXT SIMILARITY(TEXT DEDUP using IMAGE DEDUP) Link ● SCALABLE EUCLIDEAN DISTANCE Link
● SPEECH DEDUP Link ● SURVEY QUALITY Link ● FEATURE EXTRACTION Link ● COMPARE 2 TABULAR DATASET Link
● CREATE CUSTOM DATATYPE Link