

Ayushi Gaur

Data Analysis & Cleaning ► Machine Learning ► Big Data Analytics ► Computer-Vision

agaur8@uic.edu | +1 (773)-565-5914 | [LinkedIn](#) | [GitHub](#) | Taylor Street, Chicago, IL-60607

Skills & Technologies

- Programming: Python, R, OOPs using Python, Linux System Programming, Multithreading
- Data Structures: Algorithm Complexity, Arrays/Lists, Strings, Linked-Lists, Trees, Graphs, Stacks & Queues
- Exploratory Data Analysis: Pandas, NumPy, Seaborn, Matplotlib, Data Cleaning, Feature Engineering, familiar working & performing analytics on different file formats- CSV, Parquet, Excel, JSON, YAML, Config.ini, etc.
- Descriptive & Inferential Statistics- Hypothesis & Chi-Square Testing for feature selection, Central Limit Theorem, ANOVA, T-Test, Time-Series Analysis (MA, AR, ARIMA)
- Machine Learning: Linear & Logistic Regression, Cross Validation, Hyperparameter Tuning, Decision Tree, Ensemble Techniques- Bagging & Boosting (Random Forest, AdaBoost, Gradient Boosting), KNN, K-Means and Hierarchical Clustering, Feature Selection, Natural Language Processing- NLP (NER, LSTM, LDA, etc.)
- Computer Vision & Deep Learning: Image and Video Processing, CNN (LeNet, AlexNet, ResNet), TensorFlow, OpenCV, Object detection and Tracking, RCNN, Fast RCNN, YOLOV3.
- Big Data Analytics & Data Visualization: MySQL, OracleDB, PySpark, AWS - IAM, S3, Kafka, NoSQL- MongoDB & Cassandra, ETL pipelining
- Visualization Dashboards- Tableau, Azure ArcGIS for geospatial mapping, AWS QuickSight, Plotly, Dash
- Tools & Process: Docker, Git, Agile Process, Kanban

Work Experience & Projects

Magic EdTech, NY HQ
Senior ML Engineer

Big Data Analytics on EdTech Data
Feb 2022 – July 2022

Project: Designed the end-to-end ML Pipeline on Course Recommendation System on AWS platform

Problem Statement:

- The client wanted to recommend courses both online and offline to students based on their inclinations. A dedicated webpage of client was designed for the same and private APIs were exposed to download the e-books, video playlists, course details, etc. of incoming and previous students.
- The team was endowed with the responsibility to incorporate compulsory courses as a part of the curriculum along with elective courses and create an end-to-end pipeline with deployment on cloud.

Tech-Stack: Python, PySpark, Kafka, MongoDB (NoSQL), AWS- S3, EC2, ECR, GitHub, Deployment on EC2 using GitHub Actions, RASA Framework/chatbot, Machine Learning- Random Forest, Decision Tree, AdaBoost, XgBoost, etc.

Contribution:

- Data Engineering and Pipeline:
 1. Developed a Python based application to collect the data from the client's website using Rest-APIs, converted to parquet format and streamlined to KAFKA topic.
 2. Developed a consumer-based application on Linux machine in python to consume the KAFKA data and store the same in MongoDB database.
- Machine Learning Pipeline:
 1. Created an end-to-end pipeline for different ML components – Data Ingestion, Data Validation, Data Transformation, Model Creation, Model Evaluation and Model Pusher (to S3) via AWS CLI.
 2. Developed the ML models for multi-classification target and evaluated using different metrics such as Precision, Recall, F-Score, Accuracy and ROC curve. The AdaBoost performed better with accuracy and recall of 72% and this helped in proper segregation of courses for students based on their interest.
- Deployment Pipeline:
 1. Created the DOCKER image and pushed the same to AWS Ec2 and hosted the same using CI/CD pipeline tools such as GitHub Actions.

Depression analysis using Instagram posts data Research Project

IDS, University of Illinois at Chicago
September 2022 – December 2022

- **Project:** Aim is to find the difference in Instagram profiles of clinically depressed vs. non-depressed individuals

Tech-Stack: Python, R, Tableau, Git

- Scraped data from Instagram using various hashtags to find depressed and non-depressed users within a certain timeline. Images, captions and metadata features are extracted for classification •
- Performed Classification on features to identify depressed users with 63.7 % accuracy

Human Identification using Forehead Creases as a biometric Research Project

BITS Pilani, India
January 2021 – November 2021

- **Project:** Mobile based Human Identification using Forehead Creases: Application and Assessment under COVID-19 Masked Face Scenarios

Tech-Stack: Python, R, Tableau, Git

- This is the first work motivated to advance the smartphone based forehead recognition capabilities in masked face scenarios
- We presented deep metric learning using large margin cosine loss that is highly optimized for extracting discriminative features for the forehead recognition
- Incorporating a dual attention mechanism that considers spatial attention to learn semantic regularities and channel attention to compute correlation between all the channels independently.
- Devised a smart data acquisition strategy using front camera of mobile devices. The samples are acquired remotely in two sessions, our android application facilitates the process by providing onscreen GUI.
- Developed a forehead-image database, consisting of 4,940 selfie masked faces and corresponding cropped ROI's captured from 247 subjects.
- Like face recognition methods, the proposed framework has been extended to provide open-set recognition performance.
- Performance parameters: correct recognition rate (CRR) was 99.08%, equal error rate (EER) was 0.44%.

Education

Master's in Business Analytics, University of Illinois at Chicago, USA

August 2022-Present

- GPA (Fall 2022): 3.78/4.0
- Relevant Coursework: Data Mining, Deep Learning, Advance DBMS, Business Data Visualization

Ph.D. in Computer Science, BITS Pilani, India

August 2019-November 2021

B. Tech in Computer Science, Banasthali University, India

May 2014- May 2018