Shaz Ajmal

Nationality: Indian **Date of birth:** 29/03/1999 **Gender:** Male

☎ **Phone number:** (+91) 8709793086 ✉ **Email address:** shazajmal37@gmail.com

LinkedIn: https://www.linkedin.com/in/shaz-ajmal-1456b6211/

Website: https://github.com/shazam37

Home: Lane no. 1, Sattar Colony, Bariatu, 834009 Ranchi (India)

ABOUT ME

Passionate about Data Science and Machine learning, I leverage a strong background in mathematics and programming, complemented by practical expertise in end-to-end data processes, spanning collection, cleansing, and analysis. My meticulous attention to detail, coupled with a resilient work ethic and a dedication to continuous learning, has enabled me to thrive in diverse settings – from the fast-paced realm of startups, where I excelled as a Data Science intern, to the academic sphere, where I made contributions and learned extnesively as an ML research intern. Now, I am eager to apply these refined skills and insights, contributing substantively to innovative, data-centric organizations and researching on new algorithms.

WORK EXPERIENCE

ML Research Fellow

IIIT Hyderabad [08/2023 – Current]

City: Hyderabad

- Working on the generation of novel Metal Organic Framework (**MOF**) structures for CO2 adsorption energy prediction
- Implementing a **CGCNN** and **BERT** based deep learning model on CO2 adsorption energy prediction
- Exploring about different **Generative algorithms** for finding novel CO2 adsorbing MOFs.
- Using IIIT-H's Ada-**HPC** cluster for model training on remote server

Data Science Intern

EkoSight [05/2023 – 08/2023]

City: New Delhi

- Derived useful insights from Chemical data of soil using different statistical and machine learning methods
- Compiled and analysed data which helped the experimental team to optimise their chemical reactions
- Researched about soil science and analysis of data obtained from experiments (UV-vis, NIR, etc.)
- Developed an analytics dashboard using **Python**, **streamlit**, and **SQL** to store and visualise the experimental data

EDUCATION AND TRAINING

BS-MS

Indian Institute of Science Education and Research (IISER) Bhopal [2016 – 2021]

Field(s) of study: Chemistry

Thesis: Observation of Triplet State Dynamics in Organic Molecules using Fluorescence Correlation Spectroscopy

Higher Secondary

[2016]

Final grade: 94%

High School

[2014]

Final grade: 10 CGPA

SKILLS

Programming languages

- Python (Advanced)
- Javascript (Beginner)
- C/C++ (Beginner)

Libraries and Frameworks

- **Data Analysis:** Numpy, Scipy, Pandas, Matplotlib, Seaborn
- **Chemoinformatics:** RDkit, OpenBabel
- **Machine Learning:** Scikit-Learn, PyTorch, Keras, Jax, Huggingface, Langchain
- **Web scraping:** BeautifulSoup, Selenium
- **Web development:** HTML,CSS, Flask, Bootstrap, JQuery, Streamlit, FastAPI
- **App development:** Tkinter

DBMS

- MySQL
- Pinecone

Cloud Platforms

- AWS
- Azure
- Heroku

Other Tools

- **Version-Control:** Git, Github, DVC, DagsHub
- **Container tools:** Docker
- **Experiment tools:** MLflow
- **CI/CD tools:** Github-actions, CircleCI, ZenML
- High Performance Computing (Linux)
- Origin-Pro
- LaTeX
- MS-Excel

PROJECTS

Prediction of CO2 adsorption in Metal Organic Frameworks (MOF)

[08/2023 – Current]

- Processed and analysed the DFT based OpenDAC dataset (released by Meta) for CO2 adsorption energy prediction
- Wrote several Python scripts for visual inspection of different MOF structures.
- Created file conversion pipeline to load the data into the MOFTransformer we are using

- Currently working on Improving the adsorption prediction by fine tuning the model.
- Further plan on implementing different state of the art generative models
- **Link:** **https://github.com/shazam37/MOF-Generator-**

DS/ML learning projects in Chemistry

Implemented:

- **SchNet** (a self-attention **GCNN** based model) to predict the space groups of different crystal structures
- **XAI** model (using Shapley values, Integrated gradients, and Counterfactuals) that can provide actionable and complete explanation of the model for predicting the hemolytic activity of a peptide sequence
- **VAE** model for generating new molecules and regressing on the latent space of VAE for specific property prediction
- Generative **RNN** model for molecular generation that can be hosted on a browser using Javascript
- Graph Neural Network (**GNN**) model to predict the DFT energies on QM9 dataset
- Equivariant Neural Network (**ENN**) to predict the molecular trajectories

- **CNN** based classification model (on tokenized amino acids) to predict the solubility of a protein structure
- Regression models (Linear regression, Kernel based, and RNN) to predict the solubility of molecules on AqSol database
- MLP based classification model to predict whether a drug will pass the clinical trials on MoleculeNet database
- **Link: https://github.com/shazam37/DL_Chem_Projects/tree/main**

Crop Recommendation System

- Developed a model that lets user input the value of 7 primary features crucial to soil health and suggests the crop the farmer should grow, and the intervention measure they should take to improve their soil quality
- Trained the model on a comprehensive dataset scraped from web. Different classification ML algorithms were tested (Logistic Regression, LDA, KNN, Decision Tree, Random Forest, SVC, GaussianNB, Adaboost, Gradientboost). **Gaussian NB** emerged out to be the most accurate (Training Score: 99.5%, Validation Score: 99.3%)
- Users can specify the values of 7 input features (Nitrogen, Phosphorous, Potassium, Temperature, Humidity, pH, Rainfall). The model on the backend does the prediction and classification. Based on the classification result, suggestions are given.
- Built the application using **Streamlit** ready to be deployed on **Heroku**
- **Link: https://github.com/shazam37/Crop-Recommender**

Waste Object Detection

- Images of 13 different types of waste objects were collected and labelled with **autodistill**. The goal was to train a model that could identify and label the images of waste objects
- Applied image augmentation techniques. Fine-tuned **YOLOv5** (a CNN based object detection model) for the task and obtained an accuracy of 95%
- Built the CI/CD pipeline using **Github-actions**
- Built the application using **Flask** ready to be deployed on **AWS** with **Docker**
- **Link: https://github.com/shazam37/Waste_detection**

Medical Chatbot

- The goal of the project was to train a LLM on a medical textbook data
- Used **Llama-2** (a LLM model) and **Langchain** to create a pipeline for taking user query (a medical term or disease) and generating a response. The vector data was stored and retrieved using **Pinecone vector DB**
- Built the CI/CD pipeline using **Github-actions**
- Built the application using **Flask** ready to be deployed on **AWS** with **Docker**
- **Link: https://github.com/shazam37/Medical_chatbot**

Text Summarizer

- The data was obtained from HuggingFace. Contains dialogue between people and their summaries. The goal was to train a model that summarises a dialogue.
- Fine-tuned Google's **PEGASUS** (a Generative LLM) on the dataset to generate summary with respect to a given dialogue. The model's performance was measured through Rouge Score (0-1, 1 being best). Obtained 0.6 score on Rouge2 metric.
- Built the CI/CD pipeline using **Github-Actions.**
- Built the application using **FastAPI** ready to be deployed on **Azure** with **Docker**
- **Link: https://github.com/shazam37/Text_Summarizer-Project**

Kidney Disease Classifier

- The data was obtained from Kaggle. Contains the **CT-Scan Images** of kidney with respect to 4 labels: normal, stone, cyst, tumour. The goal was to classify the image.
- Fine-tuned the data on **VGG-16** model to predict the status of kidney from its CT-Scan Image. Obtained test accuracy of 80%. Exclusive consideration was given to False Negatives.
- Built the CI/CD pipeline using **Github-Actions**. Experimented with different hyperparameters using **MLFlow** and kept track of the model artifacts using **DagsHub** and **DVC**
- Built the application using **Flask** ready to be deployed on **AWS** with **Docker**
- **Link: https://github.com/shazam37/Kidney_Disease_Classification**

Hate Speech Classification

- The data was obtained from Kaggle (Twitter Hate Speech). The goal was to categorise a text sentence as Hateful or No Hateful.
- Clean and pre-processed the data using standard NLP methods. The data ingestion was done through **GCP B ucket**
- Developed a custom **LSTM** model for training and got the accuracy around 95%
- Built the CI/CD pipeline using **CircleCI**
- Built the application using **FastAPI** ready to be deployed on **AWS** with **Docker**
- **Link: https://github.com/shazam37/Hate-Speech-Classification**

Book Recommender System

- The data of the books, users, and reviews were obtained from Kaggle. The goal was to build a recommender system using collaborative filtering.
- Done proper data analysis utilising merge, group-by, and pivot table operations. Used **KNN** for recommending the similar books together
- Built a book recommender application using **Flask** ready to be deployed on **Heroku**
- **Link: https://github.com/shazam37/Book-Recommender-System**

Student Performance Prediction

- The data was obtained from Kaggle. The goal was to predict the performance in Maths test for a student based on several other features
- Done exhaustive EDA and statistical analysis. Tested several regression machine learning models utilising Cross Validation: Random forest, Decision tree, Gradient boosting, Linear Regression, KNN, SVM, XGB, CatBoost, and AdaBoost. Got the best performance on **Linear Regression** with an R2 score of 0.88.
- Built the CI/CD pipeline using **Github-actions**
- Built the application using **Flask** ready to be deployed on **AWS** with **Docker**
- **Link: https://github.com/shazam37/ML_Project/tree/main**

Customer Satisfaction Prediction

- The data was obtained from Kaggle, compiled, cleaned and analyzed. The goal was binary classification of a customers response.
- Tested different regression models. (Still testing). So far obtained the best R2 score of 0.7 on RandomForest Regression.
- Built the CI/CD pipeline using **ZenML** and integrated it with **MLFlow** for keeping track of the experiments.
- Built the application using **Streamlit**
- **Link: https://github.com/shazam37/Customer_Satisfaction_Predictor**

Python learning Projects

- **Auto Birthday-Wisher-** developed a code that can be hosted on the cloud to send birthday wishes through emails
- **Snake game-** developed the popular snake game using Tkinter
- **Coffee-Machine-** built an application that replicates the work of a standard coffee machine
- **Password manager App-** built an application that keeps a track of the passwords used in different places and can generate a new one
- **Quiz App-** built an application of a quiz game
- **Stock-market news-** developed a code that tracks the companies of interest and alerts about the status of its stock prices through SMS
- **Flight deal tracker-** built an interface that informs about the cheapest flight deals from a given city to other locations. Anyone can register for the service by providing their email ids
- **Cookie Manager-** automated the process of playing cookie manager game on the internet using Selenium
- **Website-** Developed a simple blog website using Bootstrap and Flask
- **Link: https://github.com/shazam37/My_Python_Projects/tree/main**

HONOURS AND AWARDS

List of titles and honours:

- **DST INSPIRE Scholar** (Issued by DST, Govt. of India) for pursuing pure Science and scoring in the top 1% percentile in 12th Boards
- Runner-up in **Civic Tech Hackathon** (Issued by CII.CO, IIM Ahmedabad) for pitching a social startup idea. Received incubation and funding for 6 months where we tried testing and scaling our idea
- **Editor/Writer** for a book summary startup (2021-22) where I guided 10 interns and summarised more than 100 books
- Subject Matter Expert in chemistry at **Chegg** (2022-2023) where I helped solve more than 4000 questions
- **GATE** (Chemistry) Qualified: secured 1800 rank
- **JEE Advanced** Qualified: secured 12000 rank

Roles and Responsibilities

- Served as a member of School Debate Society
- Served as a member of School Quiz Team winning intra and inter-school competitions
- Served as member of Institute's Chemistry Club
- Served as a batch and mess representative for the year 2017-18
- Served as a member of Institute's Students Activity Council
- Served as a member of Institute's Book Club
- Actively participated in Rural Development Initiative (Swacch Bharat Abhiyan)

HOBBIES AND INTERESTS

List of Hobbies:

- wide ranging reading interest from Philosophy, Sci-fi, Psychology, History to Neuroscience
- Learning new songs on Guitar
- Cooking and Baking occasionally
- Actively play Chess