

Gaurav Jain

+91-7011579022 • gj979986@gmail.com • [GitHub](#) • [LinkedIn](#)

PROFESSIONAL SUMMARY

A curious and passionate individual with a profound interest in research, mathematics, data science, astronomy, genomics, and programming. Enthusiastic about the potential of data to drive informed decisions, particularly in the realms of machine learning and data science. Constantly seeking opportunities to apply and experiment with various algorithms and techniques to tackle real-world challenges. Eager to explore new avenues and embrace fresh challenges in the field.

WORK EXPERIENCE

Intern Data Scientist – Tech Mahindra, Noida

June 2023 – July 2023

- Contributed to the development of IndicLLM, a Large Language Model for Indian Languages.
- Specialized in data collection and pre-processing, utilizing Python scripts for scraping Hindi data from websites.
- Conducted extensive research to identify suitable data sources for the project.
- Implemented database population based on data type using ORM.

Intern Data Analyst – Brillanz Software Solutions LLP, Pune

June 2023 – July 2023

- Collaborated on the development of the KPI Analytics App, an analytic engine for data-driven decision-making.
- Designed the app's high-level structure.
- Implemented user authentication using Flutter and Firebase.
- Developed backend functionalities for Import Data and Upload Image features using Django.

PUBLICATIONS

Jain, G., Kumar, P., & Kumar, C. (2023). "Exploring Smartphone Driven Indoor Localization System: A Systematic Literature Review." In Proceedings of the 5th International Conference on Advances in AI for Biomedical Instrumentation Electronics and Computing (ICABEC-2023). Paper ID: RP-ICABEC-2023/43.

PROJECTS

Drug Discovery Using EDA and ML – Bioactivity Prediction Web App - [GitHub](#) | [Output](#)

Jun 2022

Research project: Investigating the role of fingerprints in explaining variations in IC50 values of compounds inhibiting the Human AChE enzyme using bioinformatics, exploratory data analysis (EDA), and regression analysis (ML concept). Building a web app to predict IC50 values based on the findings.

Customized Deep Learning Model for Alzheimer's Detection from MRI Images - [GitHub](#)

Mar 2023

Project Aim: Developing a customized deep learning model to detect four classes of Alzheimer's disease from MRI images. The objective is to enhance early detection and diagnosis of Alzheimer's disease through accessible e imaging technology.

Toxic Comment Classifier using NLP and ML - [GitHub](#) | [App](#) | [Output](#)

Aug 2022

Project Aim: Utilizing concepts of Natural Language Processing (NLP) and Machine Learning (ML) to classify English comments into six different categories. The classification will also include neutral cases, with the goal of accurately categorizing and understanding the sentiment and content of the comments.

House Price Predictor Web App – [GitHub](#) | [App](#) | [Output](#)

Apr 2022

Project Aim: Developing a web app for predicting Real Estate property prices in Bangalore. This project encompasses the entire Data Science life cycle, including data collection and preprocessing. It utilizes features such as area type, balcony, location, total_sqft, bathroom, and BHK to train a machine learning model.

EDUCATION

EDUCATION	INSITUATION	YEAR	GRADE
Bachelors in computer science & engineering with Specialization in Data Science	Amity University, Noida	2020-2024	8.45 (CGPA)
XII th Education (PCM+CS)	National Victor Public School, I.P. Extension, Delhi	2019-2020	89.8%
X th Education	National Victor Public School, I.P. Extension, Delhi	2017-2018	91.8%

SKILLS

- **Languages:** Python, C, C++, R, SQL, Flutter
- **Tools:** PyCharm, VS Code, SQL Server Management Studio, PowerBI, Android Studio, Jupyter-lab, Firebase
- **Data Analysis and Data Science:** Concepts of statistics, Data Science & its methodologies, Exploratory Data Analysis
- **Machine Learning:** Supervised and unsupervised learning, model evaluation, feature engineering
- **Deep Learning:** Neural networks, convolutional neural networks (CNNs), Artificial neural networks (ANNs), TensorFlow, Keras

CERTIFICATIONS

- **Structured Query Language** Aug 2022
Kaggle
- **R Essentials** Aug 2022
IBM – Cognitive Classes
- **Introduction to Genomic Technologies** Jun 2022
The Johns Hopkins University – Coursera
- **Applied Data Science with Python – Level 2** Feb 2022
IBM – Cognitive Classes

HOBBIES/INTERESTS

- Reading and Watching Comics (Marvel & DC)
- Play and follow football news, teams, and players.
- PC Gaming (Genre: Storyline games)
- Explore new developments happening in the tech world.