Mohan Chaturvedi

chaturvedimohan97@gmail.com | +91 9140551805 | linkedin.com/in/mohan-chaturvedi

Education

VIT Bhopal University, B.Tech in Computer Science and Engineering — Sep 2026
- **CGPA:** 7.69

City Montessori School, Indian School Certificate (ISC) — May 2022
- **Grade:** 92%

City Montessori School, Indian Certificate of Secondary Education (ICSE) — May 2020
- **Grade:** 92%

Technical Skills

Programming Languages:	Python, R, SQL
Libraries & Tools:	Pandas, Numpy, Scikit-learn, TensorFlow, Matplotlib, Seaborn, Keras, PySpark, LightGBM, XGBoost
Software & Tools:	Tableau, PowerBI, MS Excel, Docker, Azure Data Studio, Databricks, GitHub, AWS, GCP

Work Experience

Data Scientist Internship, Remote — May 2023 – Aug 2023
- Achieved 20% improvement in ML model accuracy for predicting drug solubility and blood-brain barrier penetration by implementing feature engineering and tuning hyper parameters for LGBM and XGBoost models.
- Reviewed top compound features for blood-brain barrier porosity using Pandas and Numpy for scientific testing.
- Visualized solubility trends of compounds engaging Matplotlib and Seaborn to guide team focus on drug discovery.
- Accelerated drug discovery by 80%, automating screening of 500+ millions of drug candidates through KNIME.
- Boosted scientific screening of compounds by 90%, optimizing in-house database leveraging SQL and UI design.

Projects

Neural Network-Based Plant Leaf Type and Disease Detection — July 2025
- Facilitated to a team of students to develop deep learning model for plant leaf disease detection and classification.
- Experimented 3-5-layer CNNs, ResNet50 and VGG-16 for accurate identification of leaves and diseases.
- Orchestrated cleaning of 50,000+ images of 39 plant leaves and diseases Pandas and Numpy for EDA.
- Achieved 92% accuracy employing VGG-16, aiding to advanced agricultural technology and plant health detection.

NYC Airbnb Market Analysis — February 2025
- Leveraged PySpark to find historical trends in price, location and room type across 50,000+ Airbnb's in NYC.
- Attained a 0.8 R^2 in price prediction with Linear Regression models, applying PCA and encoding techniques.
- Envisioned pricing trends across the boroughs of NYC with Seaborn and Folium, creating heatmaps.

Forecasting High-Cost Healthcare Clients — March 2024
- Accomplished 97% SVM accuracy, spearheading analysis of healthcare cost data to predict high cost clients.
- Engineered and validated features through data cleaning and EDA to enhance model performance by 32%.
- Developed decision tree model, providing actionable insights into cost drivers for better service to HMO.

Formula 1 Race Data Analysis — August 2023
- Modeled data pipeline, integrating Azure Blob Storage & Databricks to study F1 race data from 1950 to 2022.
- Executed data preprocessing using PySpark and Pandas resolving metadata issues and ensuring data integrity.
- Employed Microsoft Power BI to transform the data into compelling visualizations for deeper analysis.

Certifications

- Oracle Cloud Infrastructure 2025 Data Science Professional
- IBM Data Science Professional Certificate