Ankit Choudhary

LinkedIn • Github • Leetcode

PROFESSIONAL SUMMARY

Data Scientist with hands-on experience in data preprocessing, exploratory data analysis (EDA), feature engineering, and machine learning model development. Proficient in Python, SQL, Pandas, NumPy, Scikit-learn, and TensorFlow. Experienced in building predictive models, RESTful APIs, and deploying ML solutions with measurable performance improvements.

EDUCATION

Master of Computer Applications (Computer Applications) 2023 - 2025
J.C.Bose university of science and tech,Y.M.C.A Grade: 8.0/10.0

Bachelor of Science (Computer Applications) 2019 - 2023
J.C.Bose university of science and tech,Y.M.C.A Grade: 7.9/10.0

EXPERIENCE

Software Engineer – Data & Automation November 2025 - Present
Netmaxims pvt ltd *Faridabad, Haryana*

- Designed and automated **data validation pipelines** using Python to ensure accuracy and consistency across datasets
- Worked with **structured and semi-structured data** to identify anomalies, trends, and edge cases using Pandas and NumPy
- Developed and tested **RESTful APIs** using FastAPI and Postman for data-driven and backend services
- Performed **data quality checks, schema validation, and regression testing** for analytics workflows
- Collaborated with data engineers, backend teams, and product stakeholders to debug data-related issues
- Used **SQL** to validate records, joins, aggregations, and analytical queries
- Leveraged **Git, Jira, and CI/CD pipelines** to maintain reliable and scalable automation frameworks

Software Engineer – Data & Automation February 2025 - November 2025
BugHunters pvt ltd *Faridabad, Haryana*

- Automated data validation and test pipelines using **Python**, improving data reliability across systems
- Worked with structured datasets to analyze trends, inconsistencies, and edge cases using **Pandas and NumPy**
- Built and tested **REST APIs** using FastAPI and Postman for backend and data-driven workflows
- Assisted cross-functional teams in debugging data-related issues and improving system performance
- Used **Git and Jira** for version control, test tracking, and agile collaboration

PROJECTS

House Price Prediction (Github) August 2025
House Price Prediction — Python, Scikit-learn, Pandas, SQL
• Cleaned and preprocessed housing data with missing values and outliers
• Conducted exploratory data analysis (EDA) to identify key price-driving factors
• Built regression models including Linear Regression and Random Forest, achieving R^2 score of 0.87
• Improved prediction accuracy through feature engineering and hyperparameter tuning

Stock Market Prediction August 2025

- Performed time-series analysis and preprocessing on historical stock data
- Engineered features such as moving averages, lag variables, and volatility indicators
- Built an **LSTM-based model** achieving **lower RMSE and MAE** than baseline models
- Evaluated performance using RMSE, MAE, and visual trend comparison

SortMedic (Github) December 2024
SortMedic — Machine Learning, Python, Random Forest, KNN
• Designed end-to-end ML pipeline including data preprocessing and feature selection
• Trained classification models to predict cardiac arrest risk, achieving 91% classification accuracy
• Evaluated models using accuracy and classification metrics
• Deployed ML model via REST API with JWT-based authentication

CERTIFICATIONS

Natural Language Processing (NLP) IIT Kharakpur (Link) July 2024

- Studied the theoretical foundations of NLP, including **linguistic structure, statistical language models, text representation methods, and semantic analysis**.
- Covered topics such as **morphology, syntax, semantics, discourse, and machine translation concepts**, focusing on the principles behind modern NLP techniques and their real-world applications.

Data Structures Apna College (Link) July 2024

- Completed an in-depth DSA course covering fundamental to advanced problem-solving techniques using **Java**. The curriculum included **arrays, strings, linked lists, stacks, queues, recursion, hashing, trees, graphs, heaps, and dynamic programming**.
- Practiced over **300+ coding problems** on platforms like LeetCode and CodeStudio, focusing on **time and space complexity optimization**. Developed strong analytical and algorithmic thinking skills, enabling efficient coding solutions for real-world problems.

SKILLS

Computer Languages: Machine Learning, Python, NoSQL, SQL, SQL, Python
Data Tools: TensorFlow, PostgreSQL, NumPy, Power BI
Software Packages: Pandas, Matplotlib, Excel, MySQL, MongoDB, Excel
Soft Skills: Time management, Research, Critical Thinking
Others: AWS, Jira, Git and Github, Postman, Problem-Solving