

SARTHAK CHAWATHE

Data Scientist

Bloomington, IN | +1 (812) 837-6816 | sarthakchawathe7@gmail.com

EXPERIENCE

Data Scientist

| Jul 2024 – Present

Berkshire Hathaway

| USA

- Developed and implemented machine learning algorithms using Scikit-learn, improving predictive model accuracy by 20% across key business use cases, including customer churn prediction and sales forecasting.
- Applied Generative AI techniques to enhance model training and drive innovation in data-driven insights, achieving a 30% improvement in model accuracy and a 25% reduction in data labeling efforts.
- Led Data Cleaning, Manipulation, Feature Engineering, Modeling, Evaluation, Optimization, Testing, and Deployment efforts, optimizing the data analysis workflow and accelerating time-to-insight by 30%.
- Accelerated data analysis using Hive to retrieve the data from the Hadoop cluster and SQL to retrieve data from Redshift and explored and analyzed the customer-specific features using Spark SQL.
- Processed Feature Interaction Generation, Feature Normalization and Label Encoding with Scikit-learn preprocessing.
- Collaborated with data engineers and operation teams to implement ETL process, wrote and optimized SQL queries to perform data extraction to fit the analytical requirements.

Founding ML Engineer & Data Analyst (AI Engineering, Diffusion Models)

| May 2024 – Oct 2024

OneChimera Incorporated

| CA, USA

- Led a team of 2 UCLA interns, fine-tuned AI models for architects, including image generation, 3D model creation, and chatbot development, leveraging techniques such as prompt engineering and domain-specific customization.
- Implemented advanced web scraping methods to extract and process over 10,000 images of houses, ensuring high-quality datasets for training and analysis.
- Researched cutting-edge frameworks and tools like LangChain to build scalable data pipelines, automate workflows, and integrate AI solutions, enhancing operational efficiency by 40% and improving project turnaround times.

Graduate Teaching Assistant (Problem Solving, Data Structure & ML Algorithm)

| Aug 2023 – May 2024

Indiana University Bloomington

| IN, USA

- Orchestrated a team of 3 teaching assistants, collaborating seamlessly to provide educational support to a cohort of 50+ students, resulting in significant improvements in learning outcomes and heightened student satisfaction.
- Graded and consulted detailed feedback on over 20 assignments, ensuring timely and constructive comments that contributed to a 30% reduction in student errors and improved overall assignment quality.

Data Analyst

| Jan 2019 – Jul 2022

Aplus Datalytics

| India

- Partnered with cross-functional teams in an Agile environment, delivering data solutions using Azure and Databricks to streamline data workflows and enhance reporting capabilities.
- Led data analysis projects using Python, NumPy, and Pandas to clean, transform, and analyze large datasets, resulting in a 35% improvement in data accuracy and delivering actionable insights that informed business strategies across multiple departments.
- Conceptualized Python and SQL to extract, clean, and manipulate data from various sources, ensuring data quality and integrity.
- Coordinated with a range of machine learning algorithms and statistical models in Python to assess model accuracy, achieving an average improvement of 20% across all models.
- Performed Risk Analysis of the requirements to identify the project critical success factors and prioritize functional requirements.
- Established and maintained interactive dashboards in Power BI, presenting key metrics and trends to stakeholders, resulting in a 25% improvement in decision-making speed.
- Teamed up with data scientists to develop and deploy ML algorithms for predictive analytics and anomaly detection.

KEY SKILLS

Language:

Python, R, SQL, SAS

Packages:

NumPy, Pandas, Matplotlib, SciPy, Scikit-learn, Seaborn, TensorFlow, Keras, PyTorch, Ggplot2

AI/ML/DL:

Linear/Logistic Regression, Decision Trees, Supervised/Unsupervised Learning, Classification, SVM, Random Forests, Naive Bayes, KNN, K Means, CNN, Ensemble Methods, Gradient Boosting, XGBoost, Time Series Algorithms (ARIMA, SARIMA, Prophet), Deep Learning (RNN, LSTM, GRU), Generative AI (including GANs, VAEs), Natural Language Processing (NLP), Reinforcement Learning, Transfer Learning, Computer Vision (object detection, image recognition)

Database:

SQL Server, MySQL, PostgreSQL, MongoDB

Visualization Tools:

Tableau, Power BI, Google Power Query, Looker Studio

Cloud Platforms:

AWS, Azure, Google Cloud Platform (GCP)

Big Data Technologies:

Hadoop, Spark, Hive, Kafka

SDLC:

Agile, Scrum, Kanban, Waterfall

EDUCATION & CERTIFICATIONS

Master of Science; Data Science

| Aug 2022 – May 2024

Indiana University Bloomington, IN, USA

Bachelor of Engineering; Computer Engineering

| Aug 2015 – May 2019

K.J. Somaiya Institute of Technology, Mumbai, India