Arbaaz Chaudhari

+91 8530467691 | arbazchaudhari5110@gmail.com | https://github.com/Arbaz123-git |
https://www.linkedin.com/in/arbaaz-chaudhari-917213230/

Experienced Data Scientist with professional experience in a young startup. Currently managing a Data De-Identification project, developing backend code from scratch with Advanced Anonymization Algorithms and Complex Functionality.

Experience

SynapseHealthTech (Synapse Analytics IT Services), *Data Scientist* | Remote — *July 2024 - Present*
- Conducted market analysis and evaluated product line profitability, guiding business decisions that impacted revenue growth by 15%.
- Developed **data models, algorithms, and quality metrics**, improving reporting accuracy by 25% and addressing key business needs.
- Collaborated with cross-functional teams to integrate **data solutions**, resulting in a 20% improvement in operational efficiency.
- Presented recommendations to stakeholders and mentored a team of 5 junior data scientists, enhancing **team productivity** by 18%.
- Stayed current on AI and healthcare trends, implementing innovative solutions that increased project effectiveness by 22%.

Associate Data Scientist | Remote — *Feb 2024 - June 2024*
- Analyzed large datasets, delivering actionable insights that **increased operational efficiency** by 22%.
- Developed **predictive models**, improving decision-making accuracy by 30% and reducing costs by 15%.
- Collaborated with cross-functional teams to implement data-driven solutions, resulting in a 25% productivity boost.
- Presented findings to stakeholders through reports and visualizations, speeding up decision-making by 28%.
- Contributed to a **12% increase** in revenue and a 10% reduction in **operational costs** through data driven initiatives.

AI and Data Scientist Intern | Remote — *July 2023 - Feb 2024*
- Used **AI/ML** cutting-edge technology to solve real-world problems. resulting in a 40% reduction in process bottlenecks and a 23% increase in overall productivity across three major projects.
- Created **data visualizations and reports** that enhanced decision-making, contributing to a 25% increase in seasonal sales forecasting accuracy and a 30% improvement in the customer segmentation, leading to a 12% rise in overall conversion rates.

Yoshops, *Data Analyst Intern* | Remote — *June 2023 - July 2023*
- **Identified trends, patterns, and anomalies** in large datasets, leading to a 22% increase in accurate product recommendations and a 15% boost in cross-selling opportunities.
- Created data visualizations and reports that enhanced decision-making, contributing to a 25% increase in seasonal sales **forecasting accuracy** and a 30% improvement in the **customer segmentation**, leading to a 12% rise in overall conversion rates.

Education

7.1/10	**Post Graduation in Data Analytics**, *Imarticus*	Mumbai, India	*2022-23*
8.9/10	**Bachelor in Information Technology**, *Mumbai University*	Mumbai, India	*2018-21*

Areas of Expertise

Machine Learning	Supervised-Unsupervised ML, Regression, Classification, Clustering, Time Series, DL, Predictive Modeling
Data Analysis	Statistics, Data Visualization, Advanced Charts, MS Excel
Programming & Tools	Python (TensorFlow, Keras, Pandas, Scikit-learn, OpenCV, PyTorch), MySQL
Advanced Techniques	NLP, Computer Vision, Prompt Engineering, Chat Bot Development, Image Analytics
Artifical Intelligence	LLMs, SLM, RAG, Gen AI, LangChain, LlamaIndex
Deployment	GCP, AWS, EC2

Additional Skills

Programming	Object Oriented Programming, Data Structure and Algorithms, HTML, CSS, JavaScript
Certifications	Data Analyst Internship Certificate, Yoshops – (July-2023)	Optical Character Recognition using Python, Udemy – (Jan-2024)	Unlocking the Power of ChatGPT in Data Science: A-Z, Udemy – (Dec-2023)	Building RAG System using LlamaIndex, Analytics Vidhya – (Sept-2024)

Professional Project

Data De-Identification System — *Feb 2024 - Present*
SynapseHealthTech (Synapse Analytics IT Service)
- Developing a **software solution** designed to enable healthcare organizations to **de-identify Personal Health Information (PHI)**.
- Goal is to preserve patient privacy while maintaining the utility of data for **research, analysis and sharing purposes.**
- Built complex **anonymization algorithms** like Mondrian with K-anonymization, extending them to even more **advanced** algorithms such as L-diversity and T-closeness.
- Used **NLP** for anonymizing Protected Health Information from text and implemented various anonymization techniques on it. **Extracted entities** like diseases, chemicals, and other biomedical concepts from medical texts, leveraging specialized models **scispaCy** which trained specifically for healthcare domain
- Skills required Advance Python Programming, Object Oriented Programming, Data Manipulation, Algorithm Building, Data Analysis, Statistics, Natural Language Processing, Data Structures and Algorithms.