

ABHISHEK WALAVALKAR

abhishek.walavalkar13@gmail.com | [LinkedIn](#) | +1 (930) 333 4993 | github.com/Abhishek2019 | [LeetCode](#)

PROFESSIONAL SUMMARY

Data Scientist with over 4 years of experience developing predictive models, customer analytics, and causal inference solutions across **public sector, finance, and healthcare** domains. **Skilled in Python, SQL, ETL, and cloud platforms (AWS)**, with a **strong foundation in classification models, A/B testing, LLM and NLP**. Passionate about using data to drive strategic decisions and improve customer outcomes.

TECHNICAL SKILLS

Machine Learning & AI: Deep Learning, NLP, LLM, Predictive Modeling, Generative AI, Statistics, A/B Testing

Programming & Frameworks: R, Java, Python, Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch, LangChain, Hugging Face

Data Engineering & Pipelines: Data Preprocessing, ETL, Apache Airflow, PySpark, MLflow, Docker, DBT, Kubernetes

Cloud & Visualization: AWS (SageMaker, Lambda, S3, EC2, Glue), Azure, Databricks, Power BI, Tableau

Databases & Storage: PostgreSQL, Redshift, Snowflake, MongoDB, DynamoDB, Neo4j, NoSQL

Certifications: Microsoft: **Power BI Data Analyst Associate**, **AWS Certified Machine Learning Engineer**, **Alteryx Designer Core**

PROFESSIONAL EXPERIENCE

Data Scientist

O'Neill School of Public and Environmental Affairs

Nov 2024 – Ongoing

Bloomington, IN, US

- Built **chatbot workflows using LangChain** and **scalable data pipelines** that leveraged **LLMs** for mission statement classification and feature extraction, reducing manual labeling and **accelerating data readiness for analytics and modeling**.
- Designed and deployed **interactive dashboards using Tableau**, enabling data-driven decision-making for philanthropic organizations and internal stakeholders.
- Improved data pipeline performance by 35%** by implementing **SQL transformation scripts and robust validation checks** on 275,000+ tax records, ensuring clean, reliable inputs for downstream machine learning models and dashboards.
- Applied **causal inference** techniques, including **inverse probability weighting (IPW)** with **causalib**, to estimate the **impact of education grant funding** on county-level graduation rates, enabling data-driven program evaluation.

Data Scientist

Sutherland Global Services

Nov 2021 – Jul 2023

Mumbai, MH, India

- Developed an end-to-end **predictive modeling pipeline** to identify loan defaulters across 30, 60, and 90+ day delinquency buckets, leveraging historical repayment data and implementing LightGBM for **financial risk classification**.
- Wrote advanced **PostgreSQL** queries to transform raw quarterly loan datasets and **integrated Apache Airflow and PySpark** to automate **ETL workflows, feature engineering, statistical analysis**, and dataset preparation for scalable risk analytics.
- Built **predictive models to forecast failure risk** across 10+ components in CT and MRI machines, leveraging historical maintenance records, sensor data, and technician field reports for proactive maintenance planning.
- Achieved over **87% precision** in failure prediction, reducing unplanned downtime and demonstrating the effectiveness of real-world machine learning solutions in critical healthcare operations.

Data Analyst

Eclerx Services Ltd.

Jun 2019 – Oct 2021

Mumbai, MH, India

- Constructed a backend system for an EdTech venture on **AWS**, covering data gathering and personalized career, learning content, and job role recommendations for over 10,000 user profiles.
- Harnessed machine learning models for targeted career and course suggestions, enhancing user engagement and learning outcomes by **tailoring recommendations** to individual profiles.
- Revamped a **chat analytics** platform analyzing 12,000+ customer interactions to **extract key insights on customer** sentiment, agent performance, and sales engagement.
- Fine-Tuned an NER model to tag entities like customer issues and agent greetings**. Identified trends in complaints, promotions, and service efficiency, resulting in a 20% increase in customer satisfaction.

PERSONAL PROJECT

[Customer Churn Prediction for Telecom Data](#)

- Developed a customer churn prediction model using Orange Telecom's dataset (3,000+ customers), leveraging EDA, Chi-square tests, and feature engineering to enhance data quality.
- Addressed data imbalance using SMOTE and class-weighted models, improving recall and fairness in predictions.
- Optimized XGBoost, achieving 95.6% accuracy and 93.1% ROC AUC, outperforming Decision Trees, Random Forest, and Logistic Regression.

EDUCATION

Indiana University Bloomington

Masters (MS) in Data Science: (CGPA: 4.0 / 4.0)

Aug 2023 – May 2025

Bloomington, IN, US

University of Mumbai

Bachelors of Engineering (BE) in Computer Science

Jul 2015 – May 2019

CERTIFICATION

- AWS Certified AI Practitioner
- Microsoft Certified: Power BI Data Analyst Associate
- AWS Certified Machine Learning Engineer – Associate
- AWS Certified Data Engineer – Associate (In Progress*)