ABHISHEK WALAVALKAR

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

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: Feature Engineering, Data Preprocessing, ETL, Apache Airflow, Spark, Hadoop, Docker, Hive
Cloud & Visualization: AWS (SageMaker, Lambda, S3, EC2, Glue), Azure, Databricks, Power BI, Tableau
Databases & Storage: PostgreSQL, Redshift, Snowflake, MongoDB, DynamoDB, Neo4j, NoSQL
Certifications: Microsoft Certified: Power BI Data Analyst Associate, AWS Certified AI Practitioner

PROFESSIONAL EXPERIENCE

Data Scientist **Nov 2024 – Ongoing**
O'Neill School of Public and Environmental Affairs Bloomington, IN, US

- Applied unsupervised learning techniques including K-Means, RoBERTa, and BART to generate semantically meaningful clusters from unlabeled mission data, enabling actionable segmentation for downstream analysis.
- Automated cluster validation and chatbot workflows using LangChain and LLM integration, improving inference accuracy from IRS Form 990 filings and enabling context-aware text generation for research insights.
- Engineered AI-powered data pipelines to process and classify over 270,000 nonprofit mission statements, utilizing LLMs for advanced feature extraction and automated organizational classification, accelerating data curation efforts.
- Boosted ETL pipeline efficiency by 35% by migrating data transformation logic from Pandas to SQL, resulting in faster query execution, streamlined updates to normalized schema, and improved responsiveness for ad-hoc analytical queries.
- Conducted large-scale data integrity assessments on 275,000+ tax records using SQL-based validation logic, identifying critical gaps and improving the consistency of cross-dataset joins.

Data Scientist **Nov 2021 – Jul 2023**
Sutherland Global Services Mumbai, MH, India

- Trained machine learning and deep learning models to predict the failure of over 10 components within CT and MRI machines, utilizing historical maintenance records, sensor data, and field reports for **Philips Healthcare**.
- Implemented machine learning models such as Random Forests and Gradient Boosting Machines (LighGBM) and deep learning approaches for classification.
- Maintained a precision of over 87%, achieving predictions that facilitated maintenance and reduced equipment downtime.
- Built an end-to-end ML workflow to predict loan defaulters across delayed payment categories (30, 60, 90+ days) using historical loan data, implementing LightGBM for financial risk modeling and classification.
- Developed PostgreSQL scripts to transform quarterly raw loan data and integrated Apache Airflow to automate ETL, feature selection, statistical analysis, data visualization, and dataset splitting for risk analytics.

Data Analyst **Jun 2019 – Oct 2021**
Eclerx Services Ltd. Mumbai, MH, India

- Engineered 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.

PROJECTS

Customer Segmentation & Recommendation System

- Analyzed 400K+ transactions from a UK online retailer using RFM and clustering techniques, uncovering customer segments with a Silhouette Score of 0.43. Devised a recommendation system to suggest popular products, enhancing customer retention.

Customer Churn Prediction for Telecom Data

- Trained a customer churn prediction model using Telecom's dataset (3,000+ customers), leveraging EDA, Chi-square tests, and feature engineering to enhance data quality.
- Optimized XGBoost, achieving 95.6% accuracy and 93.1% ROC AUC, outperforming Decision Trees, Random Forest, and Logistic Regression.

EDUCATION

Indiana University Bloomington **Aug 2023 – May 2025**
Masters (MS) in Data Science: (CGPA: 4.0 / 4.0) Bloomington, IN, US

Coursework: *Statistics, Data Mining, Database Concepts and Technologies, Machine Learning, Image Processing, Big Data Applications, Scientific Visualization*

University of Mumbai **Jul 2015 – May 2019**
Bachelors of Engineering (BE) in Computer Science Mumbai, MH, India