

Mihir Thakur

602-576-4346 • Tempe, AZ (Open to relocate) • tmihir27@gmail.com • linkedin.com/in/mihirthakur7 • mik-27.github.io

EDUCATION

Master of Science, Computer Science

Arizona State University, Tempe, AZ

May 2025

3.90 GPA

Relevant Coursework: NLP, Statistical ML, Data Visualization, ML Acceleration, Data Mining

Bachelor of Engineering, Computer Science Engineering

University of Mumbai, Mumbai, India

May 2022

3.85 GPA

PROFESSIONAL EXPERIENCE

Ascend - AI Engineer - Remote, US

June 2025 - October 2025

- Implemented 10 tools using LangChain and **PostgreSQL** on Supabase database through AI Agent tool calling and structured prompting over Temporal workflows, improving tool calling accuracy by **25%**.
- Developed **document retrieval** from OpenAI vector store, augmenting AI Agent with document context within **6 seconds**.
- Resolved issues in the Temporal workflow and AI Agent to optimize process, reducing latency by **200ms**.

LTIMindtree - Data Scientist/Data Engineer - Pune, India

June 2022 - July 2023

- Implemented 3 data science solutions using Pandas, PySpark, Scikit-learn and SQL in **Databricks**, articulating problems into data-driven insights through statistical analysis for business decision-making.
- Developed scalable **ETL pipelines** on Spark processing terabytes of data and trained tree-based models (XGBoost/LGBM) models to segment customers for targeted marketing, leading to estimated **\$1M impact** on boosting revenue.
- Identified and optimized **SQL** queries using SparkSQL ensuring data quality and validity, reducing **runtime by 20%**.
- Innovated data-driven approach to classify delinquent loan customers for risk-based loan recovery by building **end-to-end machine learning workflows**, employing MLflow for ML lifecycle management, and improving recall by relative **10%**.
- Benchmarked and evaluated new delinquency prediction model against the production baseline in controlled **A/B test**, reducing false negatives by 9% and delivering **\$1.5M reduction** in credit losses.
- Communicated model results and data-driven recommendations to business stakeholders and leadership through presentations and **PowerBI** dashboards, influencing business strategies and informing key decisions.
- Managed end-to-end lifecycle of over **15 ML solutions**, including deployment to Azure, maintenance, and troubleshooting of solution in production to ensure high availability and performance.

PROJECTS

M5 Sales Forecasting - Python, Pandas, Scikit-learn, PyTorch, SciPy, Plotly

- Engineered and benchmarked multiple **time-series techniques** (e.g., Prophet, SARIMAX, ARIMA, LSTM) for end-to-end sales forecasting prototype, achieving 12% accuracy improvement over the baseline model.

Voice Analytics - Python, Azure Cognitive Services, Azure Data Factory, Databricks

- Designed and automated end-to-end **ETL workflow** using Azure Data Factory for ingesting, transforming, and storing structured (VerticaDB) and unstructured (audio) data into a centralized Azure SQL database.

PolicyRAG - Python, ElasticSearch, HuggingFace, Flask, Gemma 2, RAG

- Engineered RAG solution by generating embeddings employing BGE-large model from 1000+ documents, employing hybrid retriever with ElasticSearch, enhancing context-awareness and lowering support workload by **estimated 30%**.
- Orchestrated scalable infrastructure using containerized **Docker** images, deploying them on Kubernetes.

Resume Analyzer - Python, LangChain, FastAPI, TypeScript, Next.js, N8N, Gemini

- Built full-stack job AI application platform featuring AI resume tailoring with **feedback loop**, cover letter generation, and intelligent form autofill, reducing application time by 70%.
- Developed voice-based interview agent using Gemini Live API to conduct behavioral and technical interviews.

CERTIFICATIONS & HACKATHONS

- CERTIFICATES:** OCI Generative AI Professional Certificate, Microsoft Certified: Azure Fundamentals (ID: I277-4775)
- HACKATHONS:** Zoom Spark Finalists (Zoom Attendance), Voxel AI CV Finalists (BeastWatch – Animal Tracking)

TECHNICAL SKILLS

Languages and Databases: Python, C/C++, JavaScript, MS SQL, PostgreSQL, MongoDB, Pinecone

Cloud: Databricks, Azure (AzureML, Blob Storage, Data Factory), AWS (S3, EC2, Sagemaker, Lambda), Docker, Temporal

Development: Numpy, Pandas, Scikit-learn, Plotly, PySpark, PyTorch, LangChain, HuggingFace, SHAP, NLTK

AI/ML: Statistics, Data Modeling, Regression, Clustering, Neural Networks, Model Evaluation, Hypothesis Testing

Others: Git, GitHub, Linux, Apache Spark, PowerBI, Jupyter, REST API, DevOps, Copilot, Jira