Mihir Borkar

Sunnyvale, CA | mihirborkar2004@gmail.com | 669-264-8602 | www.linkedin.com/in/mihir-borkar-3304641b0/ | github.com/mihirb1

Skills

Languages: Python, SQL, R, Java, C++, Bash, JavaScript, HTML, CSS
Frameworks/Libraries: Pandas, NumPy, Scikit-learn, TensorFlow, Matplotlib, Seaborn, PyTorch, Flask, React, Tableau
Tools/Software: AWS (S3, EC2, Lambda, SageMaker), Airflow, Docker, GitHub Actions, Jupyter, Power BI, GCP, Jenkins, Snowflake
Concepts: Machine Learning and Data Mining, LLMs, Natural Language Processing, Time Series Forecasting, Data Structures and Algorithms, Relational Databases, Big Data Systems, Deep Learning Algorithms, Statistical Modeling and Probability

Education

University of California, Irvine, Bachelor of Science in Data Science — Graduating: June 2026
- Organizations: Hack at UCI, Data at UCI, ICS Student Council

Relevant Coursework: Probability & Statistics, Statistical Computing & EDA, Statistical Modeling, Data Mining, Machine Learning (NLP, Time Series, Deep Learning), Relational Databases, Big Data Systems, Cloud Computing, Causal Inference

Certifications: Google Data Analytics Professional Certificate, AWS Certified Data Engineer - Associate (Pursuing), Google Foundations of Cybersecurity

Experience

Research Assistant, Massachusetts Institute of Technology – Cambridge, MA — September 2022 – September 2024
- Achieved **94.29% accuracy** with fairness metrics (**0.04 individual**, **0.06 group**) using **XGBoost** with **ROC-based postprocessing** on the **German Credit dataset**
- Designed an end-to-end evaluation framework using **NSGA-II**, **Genetic Algorithms**, and **Particle Swarm Optimization** for fairness/accuracy tradeoff modeling
- Applied advanced **causal inference** and **AUC-based metrics** for robust classification benchmarking
- Published peer-reviewed findings in **MAKE Journal**: **mdpi.com/2504-4990/6/3/105**

Software Engineer Intern, Google CSSI – San Jose, CA — June 2022 – August 2022
- Created **12 interactive websites** using **Python, HTML, CSS, and JavaScript**, incorporating **Bulma** for enhanced UI
- Designed a Connect-4 website featuring a **Minimax machine learning algorithm**, allowing users to play against a computer
- Implemented a real-time multiplayer mode using **WebSocket** for server-to-client communication

Projects

SmartCart: Price Analytics Engine | *Python, MySQL, Airflow, Flask, Google Maps API, AWS Lambda, Docker*
- Created a scalable **ETL pipeline** using **Airflow and AWS Lambda**, ingesting and transforming data from **1,000+ products** into a structured **MySQL** database
- Built custom logic to identify **price anomalies** and surface actionable insights via automated dashboards and maps
- Containerized backend using **Docker**, and deployed via systemd with weekly **automated CI builds** through **Jenkins**
- Visualized location-based price differences with **Google Maps API**, optimizing for cost-effective store selection

ZotTripper: Route Optimization Visualizer | *Python, Node.js, React, Google Maps API, Pandas, NumPy*
- Cleaned and processed **350+ addresses** using geocoding and validation logic; stored structured data for downstream use
- Developed route optimization models using **2-opt, 3-opt, and Lin-Kernighan heuristics** for efficient routing
- Analyzed distance improvements via **Haversine metrics**, iteratively benchmarking against greedy baselines
- Deployed UI with **React** and plotted dynamic route visualizations through the **Google Maps SDK**

MediScanAI: Clinical QA + Recommendation System | *Python, SageMaker, S3, Lambda, Textract, React*
- Trained a **Random Forest classifier** on structured EHR data using **SageMaker**, achieving **91% classification accuracy**
- Parsed unstructured clinical documents using **Textract + Lambda**, transforming extracted text into an **S3-based structured knowledge base**
- Delivered real-time patient Q&A functionality using an **LLM-based chatbot** trained on a curated dataset of **300+ treatments**
- Integrated **LLM responses** via a **RAG pipeline** using structured data, powered by a Bedrock **AI agent (Claude 3.7)**