Sahibnoor Singh Chahal

More About Me | csahibnoor@gmail.com | linkedin.com/in/sahibnoorsingh | github.com/Sahibnoor

Professional Summary

- **Data Engineering & Architecture:** Designed scalable data pipelines using Python, Airflow, and AWS Redshift, reducing data latency by 40% and improving model retraining speed.
- **Machine Learning & Predictive Modeling:** Built and optimized predictive models (Random Forest, Gradient Boosting, Time Series) driving measurable performance and efficiency gains.
- **Experimentation & Analysis:** Accelerated model experimentation by 30% using PySpark in Databricks and implemented data-driven frameworks for power and performance monitoring.

Experience

Data Engineer (Contract) | *PRN Pharmaceuticals* Dec. 2025 – Feb. 2026

- Reconciled duplicate patient and practice records across **9 disparate acquired-practice data sources** into a consolidated NetSuite database following company acquisitions, resolving inconsistencies from manually entered sales/marketing data.
- Built a blocking-indexed fuzzy matching pipeline in Python (`rapidfuzz`) with field-level normalization (name, address, phone, email) to score and classify candidate matches at scale without brute-force pairwise comparison.
- Built a synthetic data generator with controlled duplicate injection and realistic corruption (typos, formatting noise, case variation) to create labeled ground truth, then trained a Random Forest classifier on similarity features to validate and tune matching accuracy.
- Delivered a resolved, deduplicated dataset with source IDs mapped back to consolidated partner records for downstream use.

Team Lead - Outreach & Fundraising | *Bright Mind Enrichment & Schooling* Feb. 2025 - Feb 2026

- Designed and implemented a donor tracking system using historical giving patterns and engagement likelihood scores, enabling prioritized outreach and improving follow-up consistency.
- Maintained structured datasets of corporate and institutional donors and standardized data collection and reporting workflows across a 40+ member, multi-channel fundraising operation.

Data Engineer | *GagnaPro Solutions* Dec. 2020 – Jun. 2023

- Drove **$7.5M+ in incremental revenue** by engineering churn-segmentation pipelines on **14M+ user records**, enabling hyper-targeted retention campaigns.
- **Reduced data refresh latency by 40%** and improved cross-team availability by building 20+ automated pipelines with Python, Airflow, and AWS (EC2/Redshift).
- Cut report generation time by 37% through designing scalable Snowflake data marts and optimizing complex SQL workflows.
- Boosted executive decision-making by delivering 12+ Power BI dashboards integrated with Snowflake, **reducing dashboard load times by 50%** and enabling real-time analytics.

Academic Projects

Airbnb Price Prediction and Market Insights

- Developed a scalable Gradient Boosting pricing engine (RMSE: 12.3) simulating 15% revenue gains through competitive pricing.
- Designed end-to-end workflows in Python (pandas, sklearn) with integrated feature engineering & automated reporting pipelines.
- Delivered interactive Power BI dashboards translating predictive outputs into actionable insights for host decision-making.

Skills

Machine Learning: Regression, Classification, Gradient Boosting, Random Forest, Time Series Forecasting, Model Optimization

Data Engineering: Fuzzy Matching & Entity Resolution (rapidfuzz), Data Deduplication, Blocking/Indexing for Scale, Data Cleaning & Standardization

Programming: Python (pandas, numpy, sklearn, PySpark), SQL (ETL, CTE, Window Functions), PSQL, DAX (PowerBI)

Data Systems: Databricks, AWS Redshift, Snowflake, Airflow Performance Analytics: Experimentation Frameworks, A/B & Multivariate Testing, Resource Utilization Modeling

Additional: Exploring GPU/parallel programming (CUDA, Triton, Nsight Compute) for kernel-level performance optimization

Certifications: *MITx 6.86x: ML with Python* | *IBM Data Analyst* | *IBM Machine Learning Professional*

Education

University of Massachusetts, Isenberg School of Management | *Amherst, MA* Sep. 2023 — Dec. 2024
Master of Science in Business Analytics *(GPA: 3.78)*

Bennett University | *India* Aug. 2016 — Jul. 2020
Bachelor of Technology in Computer Science *(GPA: 3.36, Specialization in AIML)*