

# JAZIL KALIM

[jazilkalim.com](http://jazilkalim.com) | [jazilkaleem@gmail.com](mailto:jazilkaleem@gmail.com) | [jazil@uchicago.edu](mailto:jazil@uchicago.edu) | [github.com/jazilkalim](https://github.com/jazilkalim) | [datacamp.com/portfolio/jazilkalim](https://datacamp.com/portfolio/jazilkalim)  
[Chicago, IL 60615](https://chicago.il.60615) | [+1\(773\) 684-7905](tel:+17736847905) | [linkedin.com/in/jazilkalim/](https://linkedin.com/in/jazilkalim/)

## Education:

University of Chicago - MS in Applied Data Science (4.0 GPA)

September 2024 – *Expected* December 2025

NED University, Pakistan - Bachelor's in Engineering (3.7 GPA)

December 2016 – December 2020

## Work Experience:

### Afiniti:

#### Senior Data Scientist (Applied AI)

June 2024 – August 2024

- Led a high-performing team of 6 data scientists at Virgin Media, a top-tier Afiniti client, to implement AI-driven agent-caller matching solutions that generated an incremental revenue of \$1.5 million in monthly revenue.
- Overseen the design and implementation of optimal AI and data engineering solutions to maximize revenue, streamline operations, and ensure efficient management of ML models, ETL pipelines, SLAs, and client inquiries which expanded the invoice to more than 50% with new business queues.
- Directed the successful migration of data infrastructure and tools; from newer MySQL versions, ETL pipeline to Airflow, R pipelines to Python, and email alerts to interactive dashboards in Power BI and Apache Superset.
- Chaired and moderated daily meetings of the internal AI team, effectively communicating results, addressing challenges, and coordinating with key stakeholders, including clients, support teams, data engineering, SID, and senior management.

#### Data Scientist – II

December 2022 – June 2024

- Employed machine learning and probabilistic modeling techniques to rank agents and segment customers, resulting in a record 50% increase in revenue. Generated a record \$0.5 million in revenue by developing agent-level churn scores using gradient boosting and binning techniques to measure agent performance for our models on lagged metric.
- Applied Bayesian statistical methods to analyze agent performance data and identify novel features, enhancing model accuracy by 15% and adding 20 new customer features.
- Conducted in-depth big data analysis using data tables to identify seasonal trends and testing new features through A/B testing, resulting in a increase in revenue by implementing targeted pipeline improvements.
- Implemented advanced alerting and dashboarding systems to enhance operational efficiency by developing an AI Dashboard that increased efficiency and responsiveness in production monitoring and issue identification for deployed AI models which resulted in 70% reduction in production tickets and issue resolution times decreased by 50% by developing a centralized Apache Superset dashboard that eliminated the need to review 26 email alerts and dozen manual python scripts.
- Project Model Staleness; Increased client invoice by \$100k by optimizing model profitability through model refresh predictor and optimization by creating a model staleness detection system utilizing t-tests, Frobenious norm, chi-square, RMSE, log-loss, correlations statistics, and A/B testing to identify when models require updates.

#### Data Scientist - I:

May 2021 – December 2022

- Updated ML modeling pipelines with the latest data, monitored production environments, and conducted root cause analyses to address operational issues and compliance.
- Troubleshoot and resolved data quality issues, production AI pipeline errors, and monitoring script problems to ensure robust model performance and facilitate feature additions from Applied AI leadership.
- Conducted exploratory data analyses to diagnose production anomalies, identify emerging trends, and detect feature outliers. Regularly presented on data science topics and internal company tools.
- Project Feature importance and stability; Enhanced model performance and stability by developing an R pipeline to shortlist 20 important features from 500 customer features using random forest, extensive preprocessing, and correlation analysis that improved model validation by 20% and reducing unstable feature-related production issues by one-third.

## Skills:

R (Dplyr, data.table, ggplot, tidyverse, rmarkdown), Python (Pandas, Numpy, Matplotlib, Seaborn, Scikit), Big data analytics, Causal Inference SQL, MATLAB, Bash, Docker, Business Intelligence, Tableau, Power BI, Machine Learning (Regression, Classification, Clustering, Recommendation Systems, Time Series Forecasting), Deep Learning (PyTorch, TensorFlow), Spark, Recommendation Systems, Data Engineering, Data Analytics, Athena, NLP, LLMs, Transformers, Cloud Computing (Azure, AWS, GCP)

## Projects:

- **Salaries prediction using GAMs:** Developed Generalized Additive regression model and deployed it on Flask server.
- **FDI Insights:** A business analytics Tableau dashboard to analyze FDI data, forecast growth, and identify investment opportunities.
- **Car Review Chatbot:** Built a Hugging Face LLM-powered NLP chatbot achieving performance in sentiment analysis, translation, and QA.
- **Stock Market Analysis:** Scraped data through web and developed a LSTM and ARIMA model to forecast stock price.
- **Recommender System:** Created movie recommender system on GCP on PySpark utilizing graph pagerank and cosine similarity.
- **Bayesian Recommender System:** Book Review through matrix completion of missing ratings using MCMC and bayes nets.