ARPITHA THIPPESWAMY

arpitha.thipeswamy@gmail.com • 437-733-1056 • Kamloops, BC
LinkedIn • GitHub • Portfolio

PROFESSIONAL SUMMARY

Data Scientist with 2.5+ years of experience in machine learning, analytics, and data engineering across healthcare, pharmaceutical, and research domains. Currently pursuing an MSc in Data Science at Thompson Rivers University (GPA 3.76), conducting NRC-funded graduate research on ensemble predictive modeling using large-scale clinical data (MIMIC-IV). Proficient in Python, SQL, Power BI, and Azure, with hands-on experience across the full data science lifecycle - from ETL pipelines and feature engineering to predictive modeling, statistical validation, and dashboard development. Skilled at translating complex business and research problems into scalable, data-driven solutions and communicating actionable insights to technical and non-technical stakeholders. Actively expanding expertise in LLMs, transformers, and RAG systems through independent development of a conversational AI extension on graduate research.

TECHNICAL SKILLS

Languages & Programming: Python (Pandas, NumPy, Scikit-learn, Matplotlib, Seaborn), PySpark (foundational), SQL (complex queries, stored procedures, aggregations), Git, GitHub, Jupyter Notebook

Machine Learning & Statistics: Random Forest, Gradient Boosting, XGBoost, Stacked Ensembles, A/B Testing, Hypothesis Testing, Statistical Validation, Feature Engineering, Model Evaluation (AUC-ROC, F1, Cross-Validation)

Data Engineering & Big Data: PostgreSQL, Microsoft SQL Server, SSIS, ETL pipeline design, raw staging & datamart architecture, OLTP/OLAP schemas, large-scale data loading, data quality auditing

Data Visualization & BI: Power BI, Power BI Service, Tableau (foundational), Excel (advanced), dashboard design, ad hoc & standard reporting ,translating insights for stakeholders at all technical levels

AI & Advanced Analytics(*currently learning)***:** LangChain, ChromaDB, RAG pipelines, LLM interfaces, transformer architecture, NLP pipelines, SHAP-driven explainability, agentic AI, prompt engineering

MLOps & Cloud(*currently learning***):** FastAPI, Docker, GitHub Actions CI/CD, model versioning, drift monitoring, Microsoft Azure (Fundamentals), Hugging Face Spaces

Methodologies & Tools: Agile/Scrum, Jira, technical documentation, cross-functional collaboration, stakeholder presentations

PROFESSIONAL EXPERIENCE

Graduate Research Assistant | Thompson Rivers University (NRC-Funded) | Kamloops, BC | *Jul 2025 – Present*

Project: Stacked Ensemble ML Models for Multi-Department Hospital LOS Prediction in Geriatric Patients | Supervisor: Dr. Piper Jackson

- Conducting NRC-funded research on hospital length-of-stay (LOS) prediction for older adults (65+) using MIMIC-IV; designed end-to-end PostgreSQL data architecture (raw staging + analytical datamart) with OLTP & OLAP schemas for high-volume data storage and downstream ML modeling.
- Wrote and optimized complex SQL queries for data extraction, transformation, feature engineering, and auditing across multiple hospital departments; ensured data accuracy and integrity prior to modeling.
- Developed and evaluated ensemble ML models (Random Forest, Gradient Boosting, Stacked Ensembles); conducted A/B-style comparative analysis of classification vs. regression formulations, evaluated using cross-validation, AUC-ROC, and F1 metrics.
- Presented research as independent poster presenter at the 2026 ISG World Conference (Vancouver); responded to audience Q&A as co-author alongside supervisor Dr. Piper Jackson.
- Currently designing the Clinical LOS Intelligence System - a conversational AI extension for natural language querying of model findings, live LOS prediction, and SHAP-driven explanations via RAG pipeline, LLM interface, and agentic orchestration.

Graduate Teaching Assistant | Department of Physical Sciences, Thompson Rivers University | Kamloops, BC | *Jan 2026 – Feb 2026*

- Delivered hands-on instruction in applied statistics in excel for a lab-based undergraduate course, guiding students through uncertainty estimation, standard deviation, confidence intervals, and relative error analysis - translating statistical concepts into practical, interpretable findings.
- Educated a diverse cohort on rigorous data validation, precision assessment, and analytical reporting - reinforcing the importance of statistical thinking and hypothesis-driven reasoning.

Data Analyst Intern | Tutort Academy | India (Remote) | *Aug 2022 – Jul 2023*

- Completed intensive ML and data analytics program covering SQL, machine learning models, deployment, and AI technologies; gained hands-on experience with core data science tools applied to practical projects

Junior Data Scientist | Aidastech (Client: QUVA Pharma, US) | Coimbatore, India | *Sep 2021 – Jul 2022*

ML & Advanced Analytics | Customer Segmentation | Probabilistic Recommendation

Enabled prioritization of the top 20% of accounts, improving sales targeting efficiency and revenue conversion. Identified **8% increase in incremental quarterly sales opportunities**, directly supporting strategic revenue growth initiatives.

- Designed and implemented an end-to-end machine learning pipeline in Python to process, clean, and transform large-scale transactional sales and customer attribute data for advanced analytics use cases.
- Engineered customer behavioral features, including a Recency–Frequency (RF) scoring model, and applied clustering techniques to segment customers into behavioral and attribute-based cohorts.
- Developed a Growth Opportunity Matrix by integrating clustering outputs with product purchase patterns to identify high-potential customer-product combinations.
- Applied the Apriori association rule mining algorithm within customer segments to build product affinity models and uncover cross-sell and up-sell opportunities.
- Built an automated recommendation engine that runs on a monthly cycle to analyze evolving sales patterns and generate actionable product recommendations.
- Designed interactive customer intelligence dashboards using Power BI to visualize purchase behavior, product mix, and personalized recommendations, enabling account managers to drive targeted sales strategies.

Data Engineering & BI

- Partnered with business users and leadership stakeholders at a US pharmaceutical client to gather data and analytics requirements, translating business needs into analytical solutions, SQL-based insights, and automated reporting outputs.
- Optimized SQL queries and stored procedures reducing report execution time and improving ETL pipeline reliability across multiple source systems, enhancing overall reporting efficiency and data accessibility.
- Designed and maintained scalable ETL pipelines using Microsoft SSIS, monitoring scheduled jobs and proactively resolving failures to ensure consistent and timely data delivery across enterprise systems.
- Developed interactive Power BI dashboards enabling self-service analytics, allowing business users to generate on-demand reports independently and improving decision-making speed and data accessibility.
- Acted as a data analytics liaison between technical and business teams, supporting end-user training, adoption, and Agile/Scrum-based delivery processes with well-documented reporting frameworks.
- ***Employee of the Quarter — Q1 2022***

EDUCATION

Master of Science, Data Science | Thompson Rivers University | Kamloops, BC | *Sep 2024 – Present*
GPA: 3.76 | Research Focus: Ensemble predictive modeling on geriatric clinical data (MIMIC-IV) | NRC-funded graduate research

Postgraduate Diploma, AI Architecture, Design & Implementation | Georgian College | Barrie, ON
Sep 2023 – Apr 2024 | *Grade: 87.59%*

Bachelor of Science, Economics, Mathematics & Statistics | Mount Carmel College | Bangalore, India | *May 2018 – Jul 2022* | *Grade: 77.37%*