ANJANA KARATHAT RAJEEV

Data Science Graduate

0478941055 | anjanakr1916@gmail.com | [Github](#) | [Linkedin](#) | [Portfolio](#) | Melbourne

SUMMARY

I am someone who's driven by the belief that skills come with responsibility. This belief drove me to move from India to Australia to pursue my Data Science Master's at RMIT and challenge myself independently. I'm passionate about applying data science to healthcare and environmental challenges. Through my volunteer work with Good Data Institute, I've experienced how powerful diverse teams can be when working toward social good. I want to keep creating positive impact on the environment and society

EXPERIENCE

Data Science Intern — Melbourne, Australia
InflaMed — 07/2024 - 11/2024

Machine Learning, Natural Language Processing, LangChain, Flask, Python, Image Classification, LLaMA

- Developed an advanced chatbot for skin condition diagnosis using LangChain to integrate LLaMA for text generation and VGG16 for image classification
- Integrated VGG16 image analysis with LLaMA text generation pretrained on verified skin condition research literature for comprehensive symptom evaluation
- Created user-friendly web interface with Flask for seamless interaction

Data Analyst Volunteer — Melbourne, Australia
Good Data Institute — 05/2025 - present

Dashboard, PowerBI, Looker studio, Time-series, Forecasting, Impact measurement, Python, SQL

- Good Data Institute, a volunteer organization where data professionals collaborate to help nonprofits solve business challenges through analytics and technology solutions.
- Currently collaborating with Mobilise, Australia's homeless services organization, to develop data-driven solutions for optimizing resource allocation and service delivery.

EDUCATION

Master of Science in Data Science — Melbourne, Australia
RMIT University, GPA 3.6/4.0 — 02/2023 - 12/2024

Master of Science in Applied Statistics and Data Analytics — Uusimaa, Finland
Aalto University - Student exchange program, GPA 3.2/4.0 — 01/2022 - 07/2022

Master of Science in Applied Statistics and Data Analytics — Tamil Nadu, India
Amrita Vishwa Vidyapeetham (Deemed University), GPA 9.19/10.0 — 06/2020 - 07/2022

Bachelor of Science in Mathematics — Tamil Nadu, India
Amrita Vishwa Vidyapeetham (Deemed University), GPA 8.83/10.0 — 06/2017 - 06/2020

PROJECTS

Homelessness Forecasting System: Resource Allocation Optimization — 2025
Time Series Analysis, Predictive Modeling, Python, Statistical Forecasting, Social Impact Analytics, Resource Planning

- Built predictive models forecasting regional homelessness patterns to optimize resource allocation for Australia's leading homeless services organization
- Measured existing program impact using internal datasets including surveys and client outcome data to establish baseline performance metrics.

Climate Impact Analysis: California Almond Industry | [Github](#) — 2025
Data Integration, Environmental Analytics, Agricultural Data, Climate Modeling, Tableau Dashboard

- Integrated disparate datasets (drought indices, agricultural statistics, wildfire records) to analyze climate change impacts on almond production
- Discovered counterintuitive pattern showing 1-3 year production increases post-wildfire due to farmers expanding acreage to compensate for losses

End-to-End Azure Data Pipeline: Brazilian E-commerce Analytics | **Github**   2025

Azure Data Factory, Databricks, Synapse Analytics, Medallion Architecture, Cloud Data Engineering, ETL Pipeline

- Built comprehensive data engineering pipeline processing Brazilian e-commerce data using Azure services, implementing Bronze-Silver-Gold medallion architecture
- Demonstrated scalable cloud data processing workflows from raw data ingestion through transformation to analytics-ready datasets using industry best practices

Real-time Taxi Data Stream Processing | **Github**   2024

HDFS, MapReduce

- Calculated ride metrics on-the-fly while storing results in HDFS.
- Implemented check pointing and state management techniques to ensure system reliability and fault tolerance.

Predicting User Knowledge Levels with Machine Learning | **Github**   2023

Machine Learning, Decision Tree Classifier, K-Nearest Neighbors (KNN), Scifiit-Learn, Pandas, Matplotlib, Seaborn

- Developed Decision Tree and KNN models to classify user expertise in Electrical DC Machines using academic performance data, achieving 90% and 85% accuracy respectively
- Analyzed correlations between study habits and exam outcomes, proposing integration of behavioral data for enhanced predictions

Comparing Logistic Regression & Decision tree classifier in Predicting Prevalent Hypertension | **Github**   2022

Machine Learning, Logistic regression, Decision Tree Classifier, Sfilearn, Matplotlib, Seaborn, NumPy

- Developed logistic regression and decision tree models to predict hypertension
- Decision tree classifier emerged as preferred model for interpretability and nonlinear handling

SKILLS

ETL & Data Analysis: Data Pipelines , Data Ingestion, Data Manipulation, Exploratory Data Analysis (EDA)

Data Visualization & Business Intelligence (BI): Power BI, Tableau, Dashboard Design, Looker Studio, Data Storytelling, Advanced Excel

Programming: Python (Pandas, NumPy, Matplotlib/Seaborn), SQL (Advanced Queries, Window Functions), DAX queries, R (Tidyverse, ggplot2), Shell Scripting (Bash)

Cloud & Big Data Tools: Microsoft Azure, Azure Databricks, Azure Data Factory, Azure Synapse Analytics, AWS, GCP, Apache Spark (PySpark), Hadoop

Tools & Collaboration: Version Control (Git, GitHub), Collaboration Tools (Bit bucket, Confluence, Slack)

Data Science & Machine Learning: Machine Learning, AI Integration, NLP

TRAINING & CERTIFICATIONS

Datacamp Courses – Data Analyst Track
Comprehensive training covering Power BI fundamentals, DAX programming, data visualization best practices, data preparation and transformation techniques, basic to intermediate data modeling, and practical applications through customer churn and HR analytics case studies.