Mayank Kantharia

Melbourne | +61 411683252 | mayankkantharia01@gmail.com | GitHub | LinkedIn | Portfolio

Profile

Data Scientist with a background in machine learning, NLP, and cloud-based systems. Experienced in building end-to-end analytical pipelines, deploying ML models, and developing data-driven products that support measurable business outcomes.

Key Skills

Machine Learning & Analytics: scikit-learn, PyTorch, TensorFlow, Supervised/unsupervised ML, NLP, text classification, embeddings, RAG pipelines, model evaluation, statistical analysis, feature engineering, cross-validation

Programming & Frameworks: Python, R, Flask, SQL, Next.js

AI & LLM Tools: OpenAI API (GPT-4o embeddings), vector search, semantic re-ranking

Cloud & Platforms: AWS (EC2, S3, Lambda, DynamoDB, API Gateway), REST API, FastAPI, Supabase

Databases & Tools: SQL, MongoDB, PostgreSQL, pgvector, Git/GitHub

Visualisation: Power BI, Matplotlib, Seaborn, Tableau, ggplot2

Professional Experience

Path | Data Scientist Intern **July 2025 - November 2025**

- Architected end-to-end RAG pipelines for personalised career guidance, achieving 3-second end-to-end latency and 5-millisecond vector searches, directly improving recommendation relevance and user experience
- Scaled internal role coverage by 380% (200 → 960 roles) through automated web scraping and metadata enrichment, significantly broadening the system's recommendation scope
- Designed and deployed skill-gap analysis and semantic re-ranking services via FastAPI, enabling accurate, context-aware role matching at production scale
- Contributed to full-stack delivery across ML services and frontend roadmap visualisation components within an Agile team, while implementing logging and evaluation pipelines to support continuous experimentation

Projects

Social Media & Network Analysis | Russia-Ukraine War (Reddit + YouTube) | RMIT

Python, PRAW, YouTube API, NLP, LDA, VADER, NetworkX, Louvain

- Collected and preprocessed 90,000+ comments from Reddit and YouTube, building an NLP pipeline using VADER sentiment analysis and LDA topic modelling to analyse cross-platform narrative patterns
- Detected platform-level behavioural divergence: YouTube exhibited centralised, echo-chamber-like community structures (Louvain Q=0.895, 600 communities) compared to Reddit's decentralised discourse (Q=0.571, 237 communities)
- Modelled information diffusion using the Independent Cascade algorithm, revealing YouTube's centralised hubs as amplifiers of emotionally charged, leader-focused content vs Reddit's broader geopolitical topic spread

Few-Shot Learning for Hyperspectral Data | Final Year Project | IIT Bombay

PyTorch, TensorFlow, Deep Learning, Attention Mechanisms

- Surpassed published paper benchmarks by integrating Temperature Scaling and Attention mechanisms into a Few-Shot Learning model, achieving 99.55% Overall Accuracy vs the paper's 98.16% baseline
- Designed and validated the architecture across multiple HSI benchmarks using ablation studies, confirming that Attention was the key driver - Temperature Scaling alone underperformed the baseline (97.21%)
- Performed spectral-spatial preprocessing and feature extraction on curated HSI datasets to improve model generalisation across high-dimensional class distributions

NLP Clothing Review Classification & Web Application | RMIT

TensorFlow, FastText, TF-IDF, Logistic Regression, Flask

- Engineered and benchmarked five feature representations (Bag-of-Words, TF-IDF, embeddings, title-only, combined), with the Title+Review combination achieving the highest mean accuracy of 88.97% across 5-fold cross-validation

- Trained and compared multiple classifiers (Logistic Regression, FastText, deep learning) to identify the optimal model, with combined text features consistently outperforming single-input baselines by up to 2.3%
- Deployed the best-performing model into a Flask web application supporting real-time recommendation prediction, keyword search, and user review submission

Cloud-Based Music Subscription Web Application | RMIT

AWS EC2, S3, DynamoDB, API Gateway, Lambda, Ubuntu, Apache2

- Deployed a fully functional cloud application on an EC2 Ubuntu server using Apache2, accessible via public DNS, demonstrating scalable AWS architecture
- Designed DynamoDB schemas for login, music, and subscription data with optimised key structures and built backend logic for registration, subscription management, and music queries
- Automated artist image retrieval and secure storage in S3, and integrated API Gateway with Lambda to handle user registration, subscription updates, and data retrieval via REST APIs

Education

RMIT University, Melbourne **2024-2025**
Master of Data Science (3.4/4 GPA)

Mumbai University, Mumbai **2019-2023**
Bachelor of Computer Engineering (8.96/10 GPA)