

Hem Gandhi

(Data Scientist | ML Applications & Analytics)

Melbourne, VIC • +61 484653827 • hemgandhigithub13@email.com • LinkedIn: [linkedin.com/in/hem-gandhi-92757b195](https://www.linkedin.com/in/hem-gandhi-92757b195) • GitHub: github.com/hemgandhi13 • Website: <https://hemgandhi13.github.io/>

Work Authorizations: Full Australian work rights until March 2028 (Subclass 485)

PROFESSIONAL SUMMARY

Graduate data scientist (ex-NielsenIQ, 1.3 yrs) specialising in building end-to-end data and AI applications – from recall-first churn models and geospatial visualisation to an ATS Resume Analyzer that combines classic ML, semantic similarity and LLM suggestions. Strong hands-on Python and SQL with production-minded engineering: APIs (FastAPI), geospatial stacks (PostgreSQL/PostGIS, Leaflet), Streamlit UIs, packaging and Docker. Experience spans financial forecasting/variance analysis, ATS optimisation and asset/geo data quality, with a bias towards validation, clear documentation and stakeholder-friendly communication. Actively developing as an AI Generalist/Engineer with growing depth in LLM orchestration, prompt design, vector similarity and practical MLOps habits (dependency management, containerisation, environment flags).

SKILLS

Tools / Technologies: PostgreSQL; PostGIS; relational and NoSQL databases (incl. MongoDB, document stores); Hadoop; MapReduce; XML; JSON; regular expressions; FastAPI; REST APIs; Streamlit; Leaflet.js; PyMuPDF; python-docx; Pillow; Tesseract OCR; sentence-transformers; PyTorch; LiteLLM (Claude and other configurable LLMs); Docker; Git; GitHub; Linux command line; Cloud platforms – AWS; Microsoft Azure; Oracle Cloud Infrastructure (OCI); Tableau; Excel; Excel/VBA; SPSS.

Methods / Capabilities: Problem framing and requirements capture; exploratory data analysis (EDA); data wrangling and cleaning; feature engineering; Data quality checks and validation (completeness, consistency, accuracy); statistical thinking; hypothesis testing and confidence intervals; Relational database design, ER modelling and normalisation; SQL optimisation and query tuning; basic transaction management concepts; Data modelling; ETL and data pipeline design; big data architecture and distributed processing; Classification and regression modelling (incl. recall-focused churn models); model selection; cross-validation; Model evaluation (precision, recall, F1, AUC, ROC and PR curves); class-imbalance handling (stratified splits, SMOTE, class weights); Forecasting and variance analysis; financial and scenario analysis; churn and pricing insight generation; risk indicators and KPIs; LLM and NLP applications – ATS-style scoring and resume-JD matching; keyword extraction and skills categorisation; semantic similarity and evidence ranking; Prompt design and LLM orchestration with deterministic fallbacks; retrieval/RAG and vector search fundamentals; Geospatial analysis and spatial joins; Performance optimisation; model and pipeline explainability (feature importance, SHAP); MLOps foundations – packaging, dependency management, containerised and reproducible environments with Docker; Cloud and data solution design across AWS/Azure/OCI (scalability, security, cost-awareness); Dashboarding, data visualisation and storytelling for stakeholders; Documentation, runbooks and version control best practices; Git-based team workflows (branching, pull requests, peer code review); Stakeholder communication and business-focused presentation of technical findings.

EXPERIENCE

NielsenIQ — Research Associate | Vadodara; Gujarat; India

Sep 2021 — Dec 2022

- Automated weekly client reporting for FMCG portfolios (Budweiser; L'Oréal; Dr. Oetker); lifting **on-time delivery to >98%** with **>95% accuracy**.
- Wrote SQL extracts and data checks that **reduced manual fixes ~30%** and cut the weekly cycle time.
- Produced EDA templates and variance diagnostics, cutting analysis turnaround ~25%.
- Built macro-enabled Excel templates that **standardised KPIs** across accounts and reduced rework for the team.
- Partnered with account leads to translate trends into actions (price/promo mix); **improving meeting prep and decision**

speed.

- Produced clean documentation and handover notes; improving **maintainability** across rotations.

The Maharaja Sayajirao University of Baroda; Internal Quality Assurance Cell (IQAC) — Junior Analyst (Internship) | Vadodara; Gujarat; India
Apr 2019 — Jul 2019

- Analysed feedback from ~**36,000 students** covering campus facilities; teaching quality and infrastructure.
- Built **faculty-wise** and **department-wise** summaries to support NAAC accreditation (5-year ratings) and annual planning.
- Consolidated data checks and visual summaries to improve clarity for the IQAC committee and department heads.

SELECTED PROJECTS

[ATS Resume Analyzer — JD Fit Copilot](#)

Nov 2025

- Built an **end-to-end ATS Resume Analyzer** that ingests **PDF/DOCX/TXT** resumes and job descriptions, extracts **structured skills** (tools, methods, competencies, domain) and computes simple and advanced **ATS-style scores** using **TF-IDF similarity** and **weighted keyword coverage**.
- Designed **modular parsing, scoring and evidence-ranking components** (parse, scoring, evidence modules) so the core logic can be reused in other apps/APIs; added offline heuristics plus configurable **“strict/lenient” scoring profiles**.
- Integrated an optional **semantic layer (sentence-transformers + cosine similarity)** to rank the strongest experience bullets against JD requirements and an **LLM backend (via LiteLLM)** to generate multi-dimensional fit explanations and rewrite suggestions, with **deterministic fallbacks** when keys or heavy deps are unavailable.
- Packaged and deployed as a **production-style Python app** with a **Streamlit UI**, **src/** layout, **pyproject.toml + uv** dependency groups (base vs semantic), and a lightweight **Docker image** suitable for local and cloud demos.
- **Tools & skills:** Python, pandas, scikit-learn, rapidfuzz, **Streamlit**, PyMuPDF, python-docx, Tesseract OCR, **sentence-transformers**, PyTorch (CPU), **LiteLLM**, **Docker**, Git; text parsing, ATS-style scoring, semantic similarity, LLM orchestration, packaging and deployment.

[BCG X — PowerCo Churn Analysis \(Forage\)](#)

Sept 2025

- Designed a recall-first pipeline: stratified **75/25** split; **SMOTE** on training; tree-based models; selected **XGBoost** with **threshold tuning** to align with business cost of false negatives.
- Test results: **recall ≈0.48** at threshold **0.5**; **≈0.97** at threshold **0.3** (expected precision trade-off) to prioritise **top-risk deciles**.
- Produced clear evaluation metrics and a simple plan for outreach + uplift measurement.
- Tools & skills: Python; pandas; scikit-learn/**XGBoost**; imbalanced-learn (SMOTE); Jupyter; EDA; feature engineering; threshold tuning.

[Citi FP&A — Virtual Experience \(Forage\)](#)

Sept 2025

- Built executive-style packs covering **KPI/variance analysis**; **Basel III** checks (RWA/CAR); and scenario-based **forecasting** under shock conditions.
- Reconciled 2020–2022 trends; clarified drivers; and framed base/upside/downside outlooks with clear assumptions and actions.
- Reviewed **EWT/MAT** limit signals; summarising breaches; impacts and remediation for leadership.
- Tools & skills: Excel; PowerPoint; financial analysis; scenario/sensitivity modelling; risk indicator monitoring; concise stakeholder communication.

Polimap — Dynamic Geographic Data Visualisation (NDA)

Aug 2024 — Nov 2024

- **Built ETL pipelines (Python/SQL)** and spatial models in PostGIS; exposed read-optimised FastAPI endpoints for a Leaflet front end (heatmaps; clustering).
 - Enabled responsive exploration of hundreds of thousands of rows; **heatmaps ≈ 3 s for ~ 10 k points; permission checks ≤ 50 ms.**
 - Implemented dataset-to-region RBAC, QA checks and concise runbooks; containerised services with Docker for repeatable delivery.
 - Result: awarded High Distinction. (*Work delivered under NDA.*)
 - Tools & skills: Python, SQL, PostgreSQL/PostGIS, FastAPI, Leaflet.js, Docker, Git; ETL & data modelling, spatial joins, REST API design, performance optimisation, RBAC/access control, QA, documentation.
-

ADDITIONAL EXPERIENCE

Amazon — Warehouse Associate | Melbourne; VIC

Jun 2023 — Present

- Consistently **achieved daily KPIs** (throughput; accuracy) with **zero safety incidents**; demonstrating reliability under tight SLAs.
 - Built lightweight QA routines and shift reports, reducing cycle time $\sim 20\%$ and clarifying anomalies for leads. Monitored operational KPIs and escalated risks with concise, data-first narratives.
 - Used scanners/WMS and standard work to minimise errors; supported process tidy-ups that reduced rework during peak periods.
-

EDUCATION

Performance Education — Professional Year in IT | Melbourne; VIC | Jun 2025 — Present

- Industry readiness: workplace communication; job search strategy; and Australian professional standards.

Swinburne University of Technology — Master of Data Science (Analytics) | Melbourne; VIC | Feb 2023 — Dec 2024

- Key subjects: Machine Learning; Statistical Modelling; Data Management; Cloud/APIs.
- Achievements: High Distinction for geo-analytics capstone (NDA).

The Maharaja Sayajirao University of Baroda — B.Sc. in Statistics | Vadodara; India | 2018 — 2021

- Core areas: Probability; Inference; Regression; Time Series; Survey Methods.
-

REFERENCES

Referees available on request.