Harshmanpreet Singh

Edmonton, AB | 780-446-0505 | harshman@ualberta.ca
Git Hub: @harshman7 | www.linkedin.com/in/harshmanpreet |
https://portfolio-website-crcz.onrender.com |

EDUCATION

Bachelor of Science in Computing Science.   *Class of 2026*
University Of Alberta, Edmonton, AB.
Cumulative GPA: 3.68
Certifications & Training: *AWS Certified Solutions Architect - Associate, Google Cloud Associate Cloud Engineer.*

WORK EXPERIENCE

Data Engineer   *September 2024 - May 2025*
Disney Streaming | New York City, New York.
- Led planning, execution, and validation of cross-system data migrations, partnering with product, infrastructure, and analytics teams to minimize downtime and defects.
- Built automated ETL validation suites (schema checks, row-level reconciliations, data-quality alerts) to ensure data integrity and consistency across legacy and target platforms.
- Optimized batch and streaming data pipelines post-migration (partitioning, indexing, concurrency), improving end-to-end performance by ~30% and stabilizing SLAs.
- Authored and maintained migration runbooks, technical documentation, and architecture diagrams to speed onboarding and de-risk future migration waves.

Data Intern   *May 2025 - September 2025*
AccelerEd (via SKKY Analytics & Consulting Inc.) | Adelphi, Maryland.
- Supported data operations for higher-education systems by extracting, cleaning, and reconciling student and administrative records across PeopleSoft, OnBase, and internal tools.
- Built ad-hoc reports for enrollment and finance stakeholders, performed routine data quality checks (duplicates, missing fields, cross-system mismatches), and documented data flows and issues to improve onboarding and handover as the product transitioned back into UMGC Ventures.
- Used SQL and spreadsheet-based analysis to answer stakeholder data questions, streamline recurring reporting tasks, and surface anomalies needing business review.
- Collaborated with product and operations teams to translate business questions into repeatable queries and reporting templates, improving clarity and consistency of data requests.

Data Intern   *May 2023 - January 2024*
WFS. Transport Ltd. | Milton, Ontario.
- Analyzed large transportation and logistics datasets using Python and statistical tools to generate actionable insights for operations.
- Developed and maintained reports/dashboards to track key KPIs (route efficiency, load utilization, on-time performance) for leadership.
- Partnered closely with the company's founder, contributing data-driven input to strategic decisions in a trucking business operating for 17+ years.
- Streamlined recurring operations reporting by standardizing data pulls and templates, reducing manual effort for weekly performance reviews.

PROJECTS

DocSage (Intelligent Document Processing & Analytics) | *Python, FastAPI, Streamlit, PostgreSQL, FAISS, Ollama, Docker.*
- Built an end-to-end "insight agent" that ingests PDF documents, performs OCR, extracts structured fields, and stores them in a relational database.
- Implemented a local RAG pipeline with sentence-transformer embeddings and FAISS so the agent can ground its answers in document content.
- Designed a tool-using LLM agent (SQL, metrics, and RAG tools) that can answer questions like "which invoices are overdue and how has that changed month-over-month?" using real database queries.
- Exposed the system via FastAPI and a Streamlit UI, using Docker + docker-compose with PostgreSQL to mirror an AWS-style architecture (Textract + RDS + OpenSearch + Bedrock) at zero cost.

Whimsy (Social Media Application) | *Firebase, Java, XML, Kotlin, DALL-E.*
- Built a full-stack social platform for sharing short-form "moods," using Firebase for authentication, data storage, and robust sign-in/sign-up flows.
- Implemented social features including follow relationships, real-time feeds, user tagging, profile editing, and recommendation logic to suggest accounts to follow.
- Integrated AI features such as DALL·E-based mood image generation and automatic mood explanations triggered via the `#generate-response` hashtag.

GridSense (Formula One Outcome Predictor) | *Python, NumPy, Pandas, LightGBM, scikit-learn, SHAP, FastF1 API, Streamlit, Matplotlib, Seaborn*
- Built an end-to-end LightGBM model using FastF1 race data to predict Grand Prix winners, incorporating driver performance, team dynamics, and track-specific features (e.g., grid position impact, teammate deltas, normalized season progress).
- Improved model robustness and accuracy by engineering advanced features, imputing missing team-level data, and addressing class imbalance via threshold tuning and class weighting, with interpretability provided through SHAP analysis.
- Deployed an interactive Streamlit analytics dashboard to visualize win probabilities, driver consistency by circuit, and team win-rate breakdowns for real-time exploration.

ADDITIONAL SKILLS & INFORMATION

Technical: AWS, C, C++, Celery, Django, Docker, FastAPI, Firebase, HTML, Java, Java/Kotlin (Android), JavaScript/TypeScript (React), LangChain, LangGraph, Linux, Machine Learning (ANN, CNN), Python (Pandas, NumPy, scikit-learn, LightGBM), , Retrieval-Augmented Generation (RAG), SQL (PostgreSQL, SQLite), Streamlit, Tailwind CSS.
Developer Tools: Git, Cursor, Jira, Langsmith.
Language Proficiency: English (fluent), Punjabi (fluent), Hindi (fluent), French (beginner).
Driving License: Class 5