Simarpreet Kaur

Kaursimar9464@gmail.com | **(902)880-8691** | Simarpreet-kaur.com | LinkedIn | GitHub

EDUCATION

McMaster University – Honours Computer Science 2022 - 2026

St Lawrence College – Computer Networking and Technical Support 2017 – 2019

TECHNICAL SKILLS

Languages & Tools: Python, SQL, HTML, Git/GitHub, Jupyter, Tailwind
ML / AI: PyTorch, Sentence-Transformers (SBERT), scikit-learn, XGBoost, NLTK, semantic search, sentence embeddings, multilingual text, regex parsing, F1/ROC-AUC, cross-validation, error analysis, embedding precomputation
Data / Viz: Pandas, NumPy, Matplotlib, Seaborn, Tableau, Power BI, Excel

PROJECT EXPERIENCE

Pocket-AI
Trigger-word system that converts spoken commands into clean structured data.

- Built a real-time wake-word detection and audio-streaming pipeline that activates on a custom trigger word and captures continuous voice input with sub-500ms latency.
- Integrated Whisper AI to perform high-accuracy speech-to-text transcription, optimized for natural, noisy mic recordings collected during your dataset creation.
- Developed a robust regex-powered NLP parsing engine that converts 90% of free-form transcripts into reliable, structured JSON logs capturing task IDs, actions, events, timestamps, and assignees.

Multilingual News Explorer (NLP)
Designed a cross-lingual semantic retrieval and summarization app over multi-locale news feeds.

- Built semantic search with multilingual SBERT, enabling retrieval across 8 languages.
- Implemented extractive summarization via SBERT similarity graphs, producing concise multilingual briefs with 80% reduction in original text length.
- Deployed a Flask-based inference API (PyTorch-backed) with precomputed embeddings, supporting fast (<200ms) query responses.

Employee Retention Prediction - Salifort Motors
Predicted employee turnover using ML models to inform HR decisions and reduce costs.

- Compared Logistic Regression, Random Forest, and XGBoost using stratified validation, achieving strong predictive performance with high recall on at-risk employees.
- Analyzed 15,000+ HR records to uncover attrition drivers like department, workload, and hours
- Derived insights to support retention strategies and lower recruitment costs

WORK EXPERIENCE

Data Science Prompt Engineer – Outlier Oct 2025 - Present

- Designed and refined domain-specific prompts for data science tasks including EDA, feature engineering, statistical analysis, and ML modeling.
- Executed each task end-to-end in Python (Pandas, NumPy, scikit-learn) to produce gold-standard solutions for model evaluation.
- Analyzed gaps between human and model outputs to identify failure modes and improve numerical accuracy and analytical reliability.