

# Simarpreet Kaur

[KaurSimar9464@gmail.com](mailto:KaurSimar9464@gmail.com) | (902)880-8691 | [Simarpreet-kaur.com](http://Simarpreet-kaur.com) | [LinkedIn](#) | [GitHub](#)

---

## PROFESSIONAL SUMMARY

Data Scientist / Machine Learning Engineer with a strong foundation in Computer Science (McMaster University) and hands-on experience building end-to-end NLP systems, semantic search pipelines, and data-driven analytics workflows. Experienced in developing production-style ML pipelines involving embeddings, real-time audio processing, and statistical evaluation. Seeking roles in Data Science, Machine Learning, or AI Engineering.

---

## EDUCATION

**McMaster University** – Honours Computer Science

Graduated May 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 conversational recordings across varied acoustic environments.
- 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.

### Procurement KPI Analysis - Kaggle Dataset (SQL | Pandas | Tableau)

- Performed end-to-end EDA on a Kaggle procurement dataset using SQL and Pandas to analyze supplier performance and key operational KPIs.
  - Built Tableau dashboards to visualize procurement trends and identify cost-saving and efficiency opportunities.
- 

## WORK EXPERIENCE

### Data Science Analyst - Outlier

Oct 2025 - Present

- Designed structured evaluation inputs for model testing across data science tasks, assessing output quality, correctness, and alignment with expected analytical standards.

- Performed detailed error analysis on model outputs, identifying recurring failure patterns and recommending improvements to enhance consistency and reliability.
- Applied statistical evaluation methods (F1-score, ROC-AUC, precision/recall) to benchmark and validate model performance on structured tasks.

**Specialist → Senior Specialist - Best Buy**

Oct 2022 – Present

- Promoted to Senior Specialist; mentored team members and upheld high-performance standards in a high-volume, fast-paced environment
- Applied strong analytical thinking and client-facing communication skills advising customers on complex technical product decisions