Sandeep Sohal

✉ sandeepsohal.k@gmail.com | 📱 (647)667-4744 | 🌐 https://sandeepsohal.me/
⌨ https://github.com/sohal-sandy | in https://www.linkedin.com/in/sandeep-sohal-a58404137/

Education

Toronto Metropolitan University (formerly Ryerson University) 2020 – 2023
M.Sc. in Computer Science | Specialization in AI and ML | Link to all courses *CGPA: 3.96/4.33*

Trent University 2015 – 2017
BBA Honours Program | Specialization in Information Systems & E-Commerce | Link to all courses *Dean's Honour Roll*

University of Toronto (School of Continuing Studies) Jan. 2025 - April 2025
Big Data Management Systems and Tools| Data Science Certification Course |Course Description

Experience

Lead ML Engineer June. 2025 - Present
Squirl Signs ASL |Toronto, ON *Python, SQL, GitHub, Azure, MediaPipe Holistic, CV*

- Leading a cross-functional team of 4 developers and ML engineers in building an **MVP for a real-time ASL (American Sign Language) to English translation application**, targeting release by September 2025.
- Architecting an end-to-end deep learning pipeline leveraging computer vision and NLP to enable sentence-level ASL translation from webcam video input.
- Designed a two-stage model: (1) keypoint extraction from video using **MediaPipe and sequence modeling (BiLSTM/Transformer)** for ASL gloss prediction, and (2) fine-tuned a transformer-based language model (T5/BART) to convert gloss to fluent English text.
- Spearheaded dataset integration and preprocessing of the **2M-FLORES-ASL dataset (approx. 2,000 aligned ASL gloss-English samples)**, enabling high-quality training and evaluation.
- Implemented secure and scalable data access via **Azure Blob Storage** to streamline collaboration and model experimentation.

Research Scientist July. 2024 - Nov. 2024
Konnifel:Indian Institute of Technology-Madras |Remote *Python, SQL, GitHub, RdKit, DeepChem*

- Worked under Dr. Sanjan TP Gupta's AI Lab to study large chemical datasets and build a chemical toxicity prediction project with **GCNs (Graphical Convolutional Neural Networks) and DL models**.
- Preprocessed **500,000+** molecular structures using RdKit and DeepChem.
- Web-scraped **over a million** chemical records from PubChem and MoleculeNet, expanding dataset coverage.

Data Engineer March. 2024 - Nov. 2024
Alenka Media Inc. | Vancouver, BC | Remote *Python, SQL, GitHub, Project Management, Agile, Audio Analysis*

- Worked with a team of 5 to develop an audio classification model with **90%** accuracy, categorizing genres, tempo, and language.
- Extracted over 200,000 audio features using Librosa, PyTorch, and TensorFlow.
- Generated spectrograms, mel-frequency cepstral coefficients (MFCCs) from MP3 and M4A audio files to extract relevant features from the files and reduced data pre-processing time by **40%** using optimized pipeline architecture.

Data Analyst May. 2022 - Jul. 2024
Fiverr | Toronto, ON | Remote *Python, GitHub, WebScrapping, CV, DL, Audio Analysis*

- Established and maintained professional connections with buyers on the freelancing platform and delivered ML/DL projects.
- Automated data cleaning and pre-processing pipelines, reducing client workload by **50%**.

Marketing Data Analyst March. 2020 - Nov. 2023
Health Medica | Kitchener, ON *Python, SQL, Weka Data Mining Tool, Hootsuite, GitHub*

- Increased online engagement by **35%** through data-driven digital marketing strategies.
- Forecast sales trends with **85%** accuracy using Weka and SQL-based analysis.

Research and Personal Projects

Interactive Web Application for Detection of Skin Conditions Jan. 2025 - March. 2025
Computer Vision Project *OpenCV, CNN, PyTroch, Tensorflow*

- Built a Flask web app to classify skin conditions using a CNN with transfer learning. Containerized with Docker, deployed on AWS (ECR & EC2), and integrated CI/CD with GitHub Actions.

A Study of Machine Learning and Deep Learning Models for Fake News Classification Dec. 2021 - Jan. 2022
NLP project to identify reliable and unreliable news articles. *Python, Tensorflow,Google Colab GPU, BERT*

- Processed **1M+** news articles, optimizing dataset handling for NLP applications and achieved **94% accuracy** on test data using the BERT model.

Technologies

Languages: Python, R, SQL, HTML/CSS, Java, JavaScript
Other: Git, GitHub, MongoDB, Tableau, Power BI, SQLite, Google Cloud Platform, AWS, Azure