Sandeep Sohal

✉ sandeepsohal.k@gmail.com | 📱 (647)667-4744 | 🌐 https://sandeepsohal.me/
🐙 https://github.com/sohal-sandy | 🔗 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*

Continuing Education

University of Toronto (School of Continuing Studies)   Jan. 2025 – April. 2025

Big Data Management Systems and Tools Link to course description

- This course explores the architecture of reliable Big Data systems, highlighting differences from traditional systems and addressing challenges of large-scale data processing. It covers **NoSQL databases, tools like Spark, and cloud analytics platforms such as AWS, Azure, Google, and Snowflake**.

Experience

Research Scientist   July. 2024 - Nov. 2024

Indian Institute of Technology-Madras |Remote   *Python, SQL, GitHub, Chem Informatics Tools, 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.
- 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 over **15 repeating clients** on the freelancing platform and delivered **50+** 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*

- Currently developing an end-to-end web application to classify skin conditions using deep learning and machine learning models, expected to be completed by **March 2025**. The project leverages CNN-based models with transfer learning (such as ResNet50 and VGG-16) for accurate image classification on multiple skin datasets (DermNet, HAM10000, and ISIC Archive). It includes a FastAPI backend, a Streamlit frontend for real-time predictions, with deployment on AWS using Docker for scalability.

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, LATEX, Tensorflow,Google Colab GPU*

- A total of **6 different ML& DL models** are implemented in this project (logistic regression, decision tree, naive bayes, random forest, MLP, BERT)
- Processed **1M+** news articles, optimizing dataset handling for NLP applications and achieved **94% accuracy** on test data.

Technologies

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