Sandeep Sohal

sandeepsohal.k@gmail.com | kaur.sohal@torontomu.ca | (647)667-4744
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***
Relevant Courses: Computer Vision (Python, TensorFlow, PyTorch, scikit-learn), Deep Learning, Machine Learning (SQL, Python, Tableau, AWS Big Data Technologies, AWS Lambda, AWS EC2), Digital Image Processing (MATLAB, Python), Neural Networks

Trent University 2015 - 2017
BBA Honours Program | Specialization in Information Systems & E-Commerce | Link to all courses ***Dean's Honour Roll***
Relevant Courses: Database Management Systems (MySQL, Oracle), Project Management (PMI PMBOK, Scrum, Agile, Analysis), Human-Computer Interaction (Prototyping, Design Theory), Statistics for Economics and Management, Mathematics for Economics & Management, Multimedia & Design, Software Specifications & Development (HTML, CSS, JavaScript), Systems Analysis & Design

Experience

Research and Development Scientist July. 2024 - Nov. 2024
IIT Madras | Chennai, India | Remote *Python, SQL, GitHub, Chem Informatics Tools, RdKit, DeepChem*

- Worked under Dr. Sanjan TP Gupta's AI Lab as a research and development scientist to study large chemical datasets and build a chemical toxicity prediction project with GCNs (Graphical Convolutional Neural Networks) and DL models.
- Dr. Sanjan TP Gupta is affiliated with IIT Madras and the University of Wisconsin Madison. He is also known for his contributions in bioinformatics and cheminformatics.
- Identify chemical toxicity endpoints and preprocess molecular datasets with RdKit library.
- Analyze various chemical websites like PubChem or MoleculeNet and perform web scraping to generate a larger database for research.

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

- Studied various types of audio files to build ML/DL models that perform automatic classification of music based on genre, tempo, mood, vocal type, and language.
- Generated spectrograms, mel-frequency cepstral coefficients (MFCCs) from MP3 and M4A audio files to extract relevant features from the files.
- Used Librosa, PyTorch, Tensorflow, and Scikit Learn python libraries to perform audio analysis.

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

- Established and maintained professional connections with over 100 clients on the freelancing platform.
- Completed end-to-end ML/DL projects and presented final results.

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

- Performed digital marketing and analysis by observing online traffic data using platforms like Hootsuite.
- Used Weka data mining tool to perform data analysis and perform product sale forecasting.

Research Assistant Sep. 2022 - March. 2023
Peer Medical Foundation | Remote Work *Python, MySQL, HTML, Weka Data Mining Tool*

- Contributed to a research project that focused on advancing medical and health education globally and proposed better approaches.
- Performed data collection and analysis to support the research hypothesis with quantitative and qualitative data.

Operations Manager Oct. 2018 - Oct. 2020
Mary Brown's Franchise | Mississauga, ON *Quick Service Software, QuickBooks Accounting Software*

- Managed daily operations and performed HR duties, which included updating monthly schedules, issuing pay cheques, budgeting, and looking after the hiring and recruitment procedure.

Research Projects

Histopathological Image Classification with Pre-Trained Deep Learning Models Dec. 2021 - Feb. 2022
A multi-class medical image classification project. *Python, MATLAB, LATEX, PyTroch, Tensorflow, Scikit learn*

- The pre-trained VGG-16, ResNet-50, and Inception-V3 deep learning models are implemented in this research. A simple CNN model that is 10 layers deep is also proposed.

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)

Skills

Languages:
Python, SQL, HTML, JavaScript, CSS, MATLAB, LaTeX

Technologies & Tools:
Weka Data Mining Tool, MySQL, Sage Accounting Software, QuickBooks, WordPress, GitHub, VS Code, Google Colab, Spyder, MATLAB, AutoCAD, Tensorflow, PyTorch, Scikit-Learn, Matplotlib, NLTK, NumPy, Pandas, Django, MongoDB, Tableau, Microsoft Excel, Power BI, Snowflake, SQLite, AWS Lambda, Librosa

Certifications:
Google Cybersecurity Certificate.

Databases and SQL.