

GURKAMAL DEOL

Machine Learning Engineer

@ gdeol4@uwo.ca

🌐 www.gurkamal.com

in /gurkamal-deol-508675174

🐙 github.com/gdeol4

📞 226-688-7704

EXPERIENCE

Data Engineer

Flexport

📅 May 2021 – July 2021

📍 San Francisco, CA

- Built an ETL pipeline by restructuring and developing a legacy program into a POC.
- Migrated the entire POC codebase from python 2 to 3.
- Converted 15 Python scripts to Apache Airflow DAGs to automate the ETL process.

Data Analyst

Walmart

📅 Oct 2020 – March 2021

📍 Brampton, Canada

- Improved data integrity by working with cross-functional teams to bring over 1 million points of data to updated governance standards.
- Proactively monitored data accuracy by analyzing, cleansing, and validating large datasets.
- Successfully executed 100+ daily ad hoc data maintenance requests from across the organization.

Master Data Analyst

Loblaw Companies Limited

📅 Nov 2019 – October 2020

📍 Brampton, Canada

- Analyzed and identified gaps in enterprise data affecting digital and traditional sales using Python for more than 11 million transactions daily.
- Built dashboards identifying gaps in vendor data to support senior management decision making.
- Rewrote SAS automation scripts in Python and further automated the ETL process to cut down 30-45 minutes of daily manual work.
- Wrote complex SQL queries as part of an ETL pipelines using Tera-data and Python; speeding up daily data integrity tasks up to 50%.

Bioinformatics Data Analyst

Cyclica

📅 April 2019 – Sept 2019

📍 Toronto, Canada

- Wrote an ETL pipeline in Python utilizing Pandas, Numpy, and RDKit to perform molecular data pre-processing, sanitation, and similarity searching.
- Collected information for over 17 million molecular compounds to create a private database removing the need to rely on commercial equivalents.
- Created and maintained an internal database of 300 million rows using PostgreSQL to provide industry partners with propriety insight into drug molecules.
- Reduced database query time and increased accuracy by reducing substructure search results from 1000's of molecules to less than 100 through writing complex SQL queries using complex joins, grouping, aggregation, and nested subqueries.

EDUCATION

Master's Degree - Bioinformatics

The University of Guelph

📅 August 2019

Bachelor's Degree - Genetics

The University of Western Ontario

📅 June 2017

Certificate - Artificial Intelligence

The University of Toronto

📅 Present

SKILLS

Languages and tools

- Python, R, SQL, Bash, UNIX, Docker, PostgreSQL, Virtual Box, Github, Microsoft Azure, Circle CI, Google Compute Engine

ML and web development

- Tensorflow, PyCaret, RDKit, Streamlit, Scikit-learn, Keras, Flask, Cloud Run, Nginx, CSS/HTML

PROJECTS

Mutating Selfies - Generating novel drugs

- Implemented the STONED SELFIES algorithm as a web app, using RDKit & Streamlit.
- App shows quick novel molecule generation without a GPU using a genetic algorithm.
- link to app: bit.ly/3hZGhQ8

Frodotype Text Generator

- Built an end to end NLP model trained on fiction novels by modifying OpenAI's GPT2 model. Visualized data with Javascript
- Used Tensorflow, Google Deep Learning VM, Starlette, and Google Cloud Run to build a serverless and asynchronous web app.
- Link to app: frodotype.me

Toxic SMILES - A machine learning web app

- Converted molecules to fingerprints and trained logistic regression, knn and gradient boosting models, achieving a final accuracy of 96.4%.
- Deployed and built with RDKit, Scikit, Flask, Docker, Azure.
- Link to app: <https://toxic-smiles.onrender.com>

AutoML web app for drug discovery

- Made an interactive app showing rapid model prototyping using PyCaret, RDKit & Streamlit.
- link to app: bit.ly/3ruBSYC