GURKAMAL DEOL

Data Analyst / Bioinformatician

@ gdeol4@uwo.ca www.gurkamal.com in /gurkamal-deol-508675174 github.com/gdeol4

EXPERIENCE

Bioinformatics Data Analyst (co-op)

Cyclica

April 2019 – August 2019 Toronto, Canada

- Wrote a pipeline in python utilizing Pandas, Numpy, and RDKit to perform molecular similarity searching
- Implemented a custom ETL pipeline using Python scripts to pre-process and sanitize large data sets.
- Analyzed molecular data using various statistical measures to derive insight useful to industry partners.
- Sanitized and collected information for over 17 million molecular compounds to create a private database removing the need to rely on commercial equivalents.
- Created and maintained a database of 300 million rows using PostgreSQL to be used internally and possibly a platform feature.
- Utilized Google's Compute Engine virtual machine to build, perform machine learning tasks on, and query a large SQL database.
- Built and monitored indexes to bring down processing times from minutes to seconds to increase usability.
- Wrote complex SQL queries using complex joins, grouping, aggregation, nested subqueries to pull specific records with greater efficiency.

EDUCATION

Master's Degree - Bioinformatics

The University of Guelph

August 2019

Bachelor's Degree - Genetics

The University of Western Ontario

June 2017

Deep Learning Specialization

Coursera

August 2019 - Present

SKILLS

Languages and Tools

- Python, R, SQL, Bash, UNIX, PostgreSQL, Virtual Box, Github, Google Compute Engine

Machine Learning Frameworks

- Pytorch, Scikit-learn, Keras, Caret, glmnet

Data Analysis

- Pandas, Numpy, Tableau, Plotly, Hadoop Ecosystem

Web Development

- Flask, Bootstrap, Nginx, CSS/HTML

PROJECTS

Predicting Environmental Carcinogens

- Conducted a novel research project by training various models on the Tox21 dataset to predict whether a molecule is a carcinogen.
- Converted molecules to 4 different binary fingerprints and trained models using logistic regression, knn, and gradient boosting, achieving a final accuracy of 96.4%

Drug Decode Web App

- Created a web app which converts drug names in csv lists to their molecular structure.
- Wrote the drug name conversion processes in Python, utilizing the Pandas and Numpy libraries.
- The web app was deployed using the Flask framework and was hosted using Google's App Engine.

Examining Bank Churn Using a Geosegmentation Model

- Built a geosegmentation model examining the churn of bank customers using logistic regression.
- This case study looks at feature selection, multicollinearity, and odds ratios to provide business insight.

Writing a Perceptron Learning Algorithm

- Demonstrated how to write a perceptron learning algorithm from scratch using Python code.
- Trained the perceptron algorithm on a UCI machine learning data set to classify molecules as a proof of concept.

FaQ2Var (Fastq To Variant calling pipeline)

- A variant pipeline project for a graduate bioinformatics programming class at UofG.
- The pipeline uses fastq type files containing genomic information and returns a variant calling format (VCF) file. Intermediary .BAM and .SAM files can also be used.