



NAFIS
ABRAR



<https://nafabar.github.io>
<https://github.com/nafabar>



+1 778-929-4772
nafis.abrar@alumni.ubc.ca

EDUCATION

- 4th year • **B.Sc. in Computer Science** at University of British Columbia
Focus on Machine Learning
Notable courses: Intelligent Systems/AI, Machine Learning, Advanced Database, Computer Vision, Software Engineering, Statistical Learning. ♥ Vancouver

WORK

- Jan 2019 - Present
Scikit-learn, XGBoost, SQL, matplotlib, Spark, Flask, Docker, MS Azure • **DATA SCIENTIST** ♥ Toronto
at ScotiaBank (Artificial Intelligence and Machine Learning Team)
• Currently working in an Agile team for Scotiabank's global fraud detection AI software. Applying ensemble methods for prediction and inference.
• Performed sensitivity analysis and hyperparameter tuning to improve model performance.
• Set up multiple pipelines using Microsoft Azure and Docker to parallelize the training phase. This resulted in a speed up of 50%.
• Presented results to stakeholders and potential customers.
- Jan 2018 - Aug
Django, Pandas, Scikit-learn, PostgreSQL, HTML/CSS/JS, matplotlib, MS Azure, Github, Linux • **SCIENTIFIC SOFTWARE DEVELOPER** ♥ Vancouver
at BC Cancer Research Centre (Sohrab Shah Lab)
• Worked on a machine learning research project for integrating genomic data with imaging data of cancer cells to classify dead/alive cells. The classifier resulted in a prediction accuracy of 84%.
• Performed data analysis and implemented machine learning algorithms for cancer cell clustering problems and Microsoft hololens cell visualization app using Python libraries.
• Implemented, extended and documented python APIs and REST interfaces.
- May 2017 - Sept 2017
Django, HTML/CSS/JS • **FULL STACK WEB-DEVELOPER** ♥ Vancouver
at UBC EOSC (Earth and Ocean Sciences)
• Contributed to the backend of the UBC EOSC website by creating models, views and forms using Django.
• Exported CSV files from older Drupal7 UBC website and wrote Python scripts that automatically created objects in the new Django website using the CSV data. This resulted in loading 1000+ records in the new website.



RELEVANT PROJECTS

- June 2018 • **PIMS BC DATA SCIENCE NLP CAPSTONE PROJECT (COMM100)**
Language: Python Frameworks: Pandas, Scikit Learn, TextBlob, spaCy
• Worked in a team of 10 to determine intent and create knowledge base from live chat transcripts. The data set was provided by Comm100 which includes online chat sessions.
• The goal of the project was to cluster or correlate chat sessions and build a knowledge base in an automated way using mathematical models.
<http://workshop.bcddata.ca/2018/finalpres/comm100-slides>.
- October-Present
Personal Projects • **MACHINE LEARNING/DATA SCIENCE**
Language: Python Frameworks: Pandas, Scikit Learn
• Implemented supervised and unsupervised machine learning algorithms with Python (pandas, numpy). The following algorithms are implemented: Linear Regression, K-means, KNN, RBF-Kernels and Stochastic Gradient Descent.
• Built a sentiment analyser that extracts data from Twitter given a topic. The data from the Twitter API is then processed to give a result of how people feel about the user provided topic.
• Worked on computer vision projects such as scene recognition with bag of words, face detection in a scaled representation, local invariant features and RANSAC.



Programming Languages

- Java Competent (1 year)
Projects : Restaurant App, Advanced Calculator, DNS Server, FTP Client, Gym Database using JDBC
- C Basic (8 months)
Projects : FTP Server , x86 implementation
- Python Competent (2 years)
Projects : Machine learning algorithms, Django-UBC EOSC website, Rhoads, NLP
- SQL Competent (1 year)
Projects : GYM Database, UBC EOSC website
- Unix/bash Competent (1 year)



Web development and Design

- HTML and CSS - Competent (4 months)
Projects : UBC Eosc website, Rhoads website
- TypeScript - Basic (3 months)
Projects : Insight UBC
- Django - Proficient (8 months)
Projects : UBC EOSC website, Rhoads website



Machine Learning/Data Science

- Python - scikit-learn, Pandas, matplotlib
- PySpark*
- Xgboost*
- TensorFlow*
- Microsoft Azure

*Currently learning for work project



Dev-ops

- Github
- Docker
- Jira



Publication

- Scalable whole genome sequencing of 40,000 single cells identifies stochastic aneuploidies, genome replication states and clonal repertoires.

<https://www.biorxiv.org/content/early/2018/09/13/411058>