AFRAH SHAFQUAT

New York, NY | 857.294.7932 | afrah.shafquat@gmail.com | linkedin.com/in/afrahshafquat

RELEVANT EXPERIENCE

Postdoctoral Associate: Weill Cornell Medicine **02.2020 – Present**

- Analyzed sparse, high-dimensional single-cell RNA sequencing datasets to infer differential gene expression, RNA velocity, and differentiation lineages across

Researcher: Cornell University **08.2015 – 01.2020**

- Predicted measurement error in disease phenotypes for large datasets using Bayesian hierarchical latent variable model (github.com/afrahshafquat/phelex) and Markov Chain Monte Carlo algorithms (Gibbs Sampling, Adaptive Metropolis-Hastings)

Data Science Consultant Intern: Slalom Consulting **06.2018 – 08.2018**

- Developed machine learning approach using Random Classifier and Gradient Boosting for image classification and feature prioritization improving accuracy of model to over 90%
- Automated digital content tag prediction and classification for marketing insights using network and clustering approaches

Bioinformatics Analyst: Harvard School of Public Health **07.2013 – 07.2015**

- Predicted functional annotations of unknown proteins by designing and developing Python pipeline (huttenhower.sph.harvard.edu/ppanini)
- Analyzed and processed complex big datasets using visualization and Python scripts

Research Intern: Parkland Center for Clinical Innovation **12.2012 – 02.2013**

- Improved disease identification efficiency by developing Python module to predict corrections to misspellings in Electronic Health Records

TECHNICAL SKILLS

Machine learning: Classification (random forest, gradient boosting), regression (linear, multinomial, logistic), clustering, feature engineering and selection
Statistical methods: Regression models, principal component analysis and dimensionality reduction, hypothesis testing, ANOVA
Software and programming languages: - R, R Shiny, Python (e.g. scikit-learn, numpy, scipy, pdb, pandas), Java, VB.Net, C#, MATLAB, HTML, Windows/MAC OS/LINUX, git
Selected coursework: Machine Learning, Statistical Methods I & II, Computational Genetics and Genomics, Computational and Systems Biology, Probability and Random Variables

EDUCATION

Cornell University	*PhD. in Computational Biology (2020)*
Massachusetts Institute of Technology (MIT)	*S.B. in Biological Engineering (2013)*

SELECTED PUBLICATIONS

- **Shafquat A.**, Crystal R.G., Mezey J.G., *'Identifying novel associations in GWAS by hierarchical Bayesian latent variable detection of differentially misclassified phenotypes'* BMC Bioinformatics 21, 178 (2020). https://doi.org/10.1186/s12859-020-3387-z

- **Shafquat A.**, Mezey J.G., *'A hierarchical Bayesian latent variable model for detecting misclassified phenotypes and identifying novel associations in GWAS'* Probabilistic Modeling in Genomics, Cold Spring Harbor Laboratory (2018)S