NEELAM SINHA

Data scientist at Cancer Data Science Lab, NCI, Bethesda

Email: Neelamsinha0592@gmail.com | +12096008408
https://www.linkedin.com/in/sinhaneelam/
https://github.com/sapphiresinha

Education

University of California, Merced	Aug 2018 to Aug 2020
Masters, Computer Science and Engineering	GPA: 3.63/4.0
Kurukshetra University	Sept 2010 to June 2014
B. Tech, Computer Science and Engineering	GPA: 3.6/4.0

Skills

Programming Languages	Python, R, PHP, SQL, MySQL, Unix/Linux.
Data Analysis Skills	Data handling & preprocessing, Statistical analysis, Machine Learning (ML) Libraries: NumPy, pandas, scikit-learn, SciPy, caret. Data visualization libraries: ggplot2, forestplot, matplotlib.
Cloud Computing	Google Cloud Platform, AWS.
Data Science Skills	Supervised & unsupervised ML techniques: classification, regression, regularization, clustering, dimension reduction, A/B testing, feature engineering & selection, computer vision: feature matching algorithm.
Courses (Degree/Online)	Data Structure, DBMS, Machine Learning Engineering (Nanodegree), Data Manipulation, Statistics Fundamentals, Statistic and probability, Python programmer, Data Analytics and Statistic in R, R programming.

Experience

Data Scientist at Cancer Data Science Lab, National Cancer Institute, Bethesda (USA)   Oct 2020 to Current

Using the knowledge of data science, statistics and machine learning in the domain of clinical cancer research. My key goals involve developing biomarkers to stratify patients of cancer therapy to maximize its benefits and develop novel therapies.

Teaching Assistant at University of California, Merced, Merced (USA)   Aug 2019 to May 2020

Supervised ~20 undergraduate students, guide them with homework, lab assignments and course projects.

Intern at Cancer Data Science Lab, National Cancer Institute, Bethesda (USA)   June 2019 to Nov 2019

Learned the fundamentals of data handling, preprocessing and using big data including Next Generation Sequencing (NGS)-based genomics information and a large-scale CRISPR screen. Using this, I successfully built predictive models for response for a few cancer drugs in patients.

Software Developer (PHP) Intern at Jindal Stainless Steel, Hisar (India)   Mar 2018 to June 2018

Built a complaint portal where 'Employees of company can update/ send safety-related complains.

Data Analyst Intern at Indian Institute of Management (IIM), Lucknow, Lucknow (India)   Dec 2017 to Jan 2018

Analyzed large scale dataset to determine the factors that affect the Hotel room prices.

Publication

1. **Sinha N.,** Sinha S., .… & Ruppin E., Immune determinants of the association between tumor mutational burden and immunotherapy response across cancer types, **BioRxiv** [**Link**]

2. **Sinha N.**, Sinha S., Cheng K., … & Ruppin E., Using a recently approved tumor mutational burden biomarker to stratify patients for immunotherapy may introduce a sex bias, *JCO Precision Oncology (2021)* [**Link**]

3. Zingone A., Sinha S., ..., **Sinha N.**, … & Brid Ryan., A comprehensive map of alternative polyadenylation in African American and European American lung cancer patients. *In press at Nature Communication*.

4. Sinha S, Mitchell KA, Zingone A, Bowman E, **Sinha N**, ... & Ryan BM., Higher prevalence of homologous recombination-deficiency in lung squamous carcinoma from African Americans. *Nature Cancer (2020).* [**Link**]

Project

Object detection based on feature matching (Computer Vision project):   Mar 2020 to May 2020

To perform object detection without using deep learning, we explore and use template matching, feature matching algorithms (like: BRIEF, SURF, SIFT, ORB) and successfully detect the query image with an improved accuracy of 0.93. **Technology:** Python, Numpy, OpenCV, OpenCV-Contrib, PIL, imutils.

A multivariate model to predict diabetes diagnosis: [Link]   Dec 2019 to Jan 2020

Built a supervised model to predict diabetes diagnosis in ~800 native American patients by integrating clinical demographics e.g. BMI, Age in addition to blood glucose levels and with an Accuracy of ~0.81 and AUC of 0.80, improving it by 5% against previous state-of-the-art. **Technology:** Python, Data analysis and visualization, Supervised ML (Standard-Scaler, Random-Forest-Classifier).

Predicting drug response in cancer patient using supervised learning: June 2019 to May 2020

Built an ensemble model to predict response of cancer drugs in cancer cell lines (Patient Data) by identifying and learning on the genetic interactions of drug targets, where our current best AUC is ~0.80. **Technology:** Python, Jupiter notebook, Ensemble Learning (Random-Forest-Tree), Feature selection algorithms, Wilcoxon test, PCA.

Creating customer segments based on customer buying habits: **[Link]**   Apr 2019 to May 2019

Using the customer's weekly product buying habits of a wholesale distributor, I identified the segments of customers affected in case of change in "days of delivery per week" policy. **Technology:** Python, Unsupervised ML (PCA, K-Means, A/B test).

Finding donors for charity using Machine Learning (ML): [Link]   Mar 2019 to Apr 2019

I investigated factors that affect the likelihood of charity donations being made based on real census data. Trained, optimized and tested various supervised ML models like Adaboost, DecisionTree, and GaussianNB to predict the likelihood of donations with **an accuracy of 0.86. Technology:** Python, Supervised ML (Naive Bayes, Adaboost, Grid-SearchCV).

Customized vibration patterns for phone notification (Study project): [Link]   Jan 2019 to May 2019

With an aim to study how mobile phone users perceive the urgency level of notifications by getting vibration alerts, I built a web application, where a user can define the importance of the message/call etc. only by receiving a phone vibration. We performed a study to understand what is the upper bound of vibration patterns that a user can remember in short period of time. **Technology and Platform:** PHP, JavaScript, Visual Studio code, HTML.

Creating fast neural networks for mobile devices (Course project).   Aug. 2018 to Dec. 2018

Used Computer vision technique and DNN compression techniques to reduce the model size and computation complexity with the minimum loss in desired accuracy. **Technology and platform:** Python and TX2(GPU), Library: TensorFlow, Keras.