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 &visualization, Statistical analysis, Machine Learning (ML) Libraries: numpy, matplotlib, pandas, scikit-learn, scipy, ggplot2, caret, TensorFlow, Keras).
Statistical tool	Tableau, MS Excel.
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.
Online Courses:	Machine Learning Engineering (Nanodegree), Data Manipulation, Statistic and probability.

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. **Technology:** C, Java.

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-based genomics. Using this, I successfully built predictive models for response for a few cancer drugs in patients by using their genomic information.

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. **Technology:** PHP,HTML.

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. **Technology:** R, RStudio.

Project Experience (Selected)

FDA approval of TMB threshold to select patients for anti-PD1 treatment introduce a sex bias: — Aug 2020 to current
Analyzing patients' genomic data and clinical response, we showed that the usage of a recently approved FDA biomarker to select patients for immunotherapy have different predictive power in male vs female cancer patients. **Technology:** R, R-studio, Survival Analysis, data visualization (ggplot, forest plot).

Object detection based on feature matching (Computer Vision project): [Link] — 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 Jan 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.

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).

Publication

1. **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. *In review* at JCO Precision Oncology