

PARTHVI BHUTANI

pb1967@nyu.edu | 929-326-7945

EDUCATION

New York University

New York, NY

Candidate for Bachelor of Arts in Mathematics, Minor in Data Science (GPA: 3.91)

Expected May 2021

Relevant Coursework: Probability and Statistics for Data Science (Graduate), Introduction to Data Science (Graduate), Introduction to Machine Learning, Basic Algorithms, Data Structures, Object Oriented Programming, Introduction to Data Science, Theory of Probability, Mathematical Statistics, Database Design and Implementation, Honors Linear Algebra, Partial Differential Equations, Combinatorics, Numerical Analysis, Math Modeling, Discrete Mathematics

University of Delhi

New Delhi, India

Bachelor of Science (Honors) in Mathematics, Minor in Economics

May 2017

Relevant Coursework: Linear Algebra, Public Finance, Intro to Microeconomics, Introductory Macroeconomics, Multivariate Calculus

SKILLS

Programming: Python, Java, MATLAB, R Studio

Database Management: SQLite, MySQL, MariaDB

Machine Learning Stack: Scikit-learn, Keras, NumPy, Pandas, SciPy, Theano, TensorFlow, PyTorch, Regex, Matplotlib, Seaborn

Data Analysis Skills: Data Wrangling, Predictive Modelling, Machine Learning Algorithms, Deep Learning, Adv. MS Excel

PROFESSIONAL EXPERIENCE

HDFC ERGO General Insurance Company

Mumbai, India

Data Analyst Intern

May 2020 - Aug 2020

- Designed an unsupervised machine learning model to segment 1.1 million health insurance customers who are yet to claim insurance
- Analyzed various retail factors like net-premium, sum insured, loss-ratio, up-sell, ticket size, location-tier, portability, income, and age
- Explored various underwriting factors such as loading amount, add on cover premium, BMI, marital status, and payment mode
- Executed Hierarchical-DBSCAN clustering algorithm to profile customer segments as profitable, potential-profitable and not-profitable
- Demonstrated the most profitable 200000 customers to target and spur shift to exclusive or premium policies and encourage add-ons

NYU Courant Summer Undergraduate Research Experience Program

New York, NY

Research Intern advised by Dr. S.R. Srinivasa Varadhan (Abel Prize)

Mar 2018 - Sep 2018

- Designed Markov Chain Monte Carlo Samplers for Random Sampling of high dimensional probability distributions
- Simulated the design on various probability distributions to check computational efficiency of the algorithm used
- Compared the algorithm to Gibbs Sampling Algorithm and Metropolis-Hastings Algorithm to help fine tune the parameters
- Proved efficiency of Markov process through limiting distribution, convergence and mixing properties using mathematical analysis

NYU Courant Math Major Peer Mentor Program

New York, NY

Peer Mentor

Sep 2018 - Present

- Mentoring sophomore-level math majors with course selection, internship opportunities, job search, and campus activities
- Establishing stronger relationships with both experienced and aspiring Math Majors by sharing research ideas

ACADEMIC PROJECTS

Twitter Sentiment Analysis in lieu of COVID-19

New York, NY

Academic Project advised by Prof. Brian Dalessandro

Aug 2020 - Present

- Collected tweets from March 1, 2020, processed and analyzed over 2 million tweets and tracked user geolocation and IDs
- Created functions to assess the subjectivity and polarity of tweets to classify tweets as fact/opinion and positive/negative
- Achieved precision accuracy of 97% and recall accuracy of 98% using a Naive Bayes Classifier
- Analyzed sentiment polarity between verified and non-verified profiles and polarity over time for different states in the US
- Observed strong correlation of negative tweets with spikes of new cases in different states especially in dense regions like NY and NJ
- Investigated user mobility from state to state based on user geo-location and its correlation and impact on COVID-19 propagation

NYC Department of Health and Mental Hygiene Restaurant Ratings

New York, NY

Academic Project

May 2019 - Jul 2019

- Analyzed over 190000 restaurants across NYC and how their ratings changed over time for individual restaurants as well as big chains
- Evaluated roll rate as the probability of transitioning from one grade to another and evaluated roll rate categories over A rated places
- Reaffirmed previous analysis that an A rating remained unchanged 90% of the time especially in well-known or high-rated restaurants
- Examined individual factors-Borough, Food Type, Name, Street and zip-code to determine probability of restaurant receiving grade A
- Predicted a restaurant's health grade using all features together through Logistic Regression with a 72 % accuracy on the test set

Classification of Emails as Spam and Non-Spam

New York, NY

Course Project

Jan 2020 - Feb 2020

- Designed classification problem to predict whether an email was spam or non-spam using the text of 1000000+ emails
- Applied k-NN, Random Forest, Logistic Regression and Support Vector Machine algorithms before making final prediction
- Achieved accuracy of 93% through inspection of confusion matrix, thresholds of ROC curves and k-fold cross validation