ANURAG DUTT
700 Health Sciences Dr, Chapin - B 2022C - 667 – Stony Brook, 11790 – NY
+1 (347) 740 5056 • anurag.dutt@stonybrook.edu • https://anuragdutt.github.io
www.linkedin.com/in/anurag-dutt

EDUCATION

Stony Brook University **New York**
Master of Science, Applied Mathematics and Statistics - 3.8/4.0 *Aug, 2019–Present*
Graduate Certificate - Computational and Data Science
Relevant Coursework: Natural Language Processing, Probability and Statistics, Linear Programming, Applied Linear Algebra and Calculus, Computer Vision, Convex Optimization, Reinforcement Learning (Robotics), Analysis of Algorithms, Simulation, Theory of Database Systems, Stochastic Models, Principles of Programming Languages

Columbia University **New York**
Master of Science, Operations Research (Financial Engineering) *Sep,2017–Dec,2018*
Relevant Coursework: Continous Time Models, Deep Learning, Machine Learning, Simulation and Monte Carlo Methods, Statistical Analysis and Time-Series, Programming in C, Stochastic Optimal Control and Dynamic Programming, Optimization, Discrete Optimization, Partial Differential Equations, Real Analysis, Stochastic Methods

Birla Institute of Technology, MESRA **Ranchi**
Bachelor of Engineering with Distinction, Civil Engineering - 3.67/4.0 (WES - Conversion) (71% absolute) *Jul,2009–Apr,2013*
Relevant Coursework: Multivariable Calculus, Linear Algebra, Differential Equations, Numerical Methods, Data Structures
Honors: National Talent Search Scholarship awardee for academic excellence, State Rank – 24
Undergraduate research: Analysis of local energy losses at positive and negative steps at sub-critical open channel flows using numerical methods

RESEARCH PAPERS/PROJECTS

SMOOTH-GAN: Towards Sharp and Smooth Synthetic EHR Data Generation (Accepted to AIME-2020), Data Management and Biomedical Data Analytics Lab, SBU:
A conditional GAN architecture for generating synthetic Electronic Health Records [Link]

Transmission of Information in Social Networks (Working paper - in progress), Cognition and Decision Making Lab, SBU:
This project investigates the social transmission of memory and how such transmission shapes the memories shared among individuals as the basis for understanding how social networks influence behavior [Link]

Using Natural Language Processing and Deep Learning Techniques on 8-K reports for stock-price movement prediction (Poster - Workshop on Natural Language Processing and Computational Social Science, 2020) - LUNR Lab, SBU:
We examine significance of neural text analysis algorithms to predict changes in stock value prediction [Link]

Bottleneck detection for a microservices based cloud system using Graph Convolution Networks (Proof-of-concept), Performance Analysis of Computer Systems Lab, SBU:
We use graph convolution networks to detect and localize microservices responsible for latency spikes in services level objectives and dynmically mitigate the violations [Link]

Comparison of Adverserial techniques for synthesizing electronic health records - Data Management and Biomedical Data Analytics Lab, SBU:
We compare the efficiency of multiple Generative Adverserial Networks frameworks such as GAN, AC-GAN and WGAN for generating discrete time-series of electronic health records on the MIMIC dataset [Link]

Multi-Agent Generative Adversarial Self-Imitation Learning (Advanced graduate project):
This project implements and experiments a Generative Adversarial Self-Imitation Learning(GASIL) approach for testing the optimal convergence of Joint-Action Learners for the multiagent predator-prey environment [Link]

Dropout: A Simple Way to Prevent GANs from Overfitting:
We propose to incorporate adversarial dropout in generative multiadversarial networks [Link]

Study of network selection models for portfolio optimization:
We test three clustering algorithms in the realm of machine learning and discrete optimization to test the optimization capability of each on a sample portfolio mimicking SnP 500 [Link]

Pricing an American Option using Least Squares Monte Carlo and Simulated Annealing:
In this project we discuss how to price American option using Least Squares Monte Carlo and Simulated Annealing on multiple underlying assets [Link]

RESEARCH PAPERS - CREDITED AS RESEARCH ASSOCIATE

Stock Market participation in the aftermath of a scandal:
The paper examines the impact of stock fraud revelation on investor behavior with direct exposure to fraud vis-a-vis other stock market participants, by simulating a randomized control trial experiment [Link]

From participation to repurchase – Low-income households and micro-insurance:
The paper examines the factors that drive financial portfolio choices of low-income households by analyzing repurchase patterns of micro-insurance and pension policies [Link]

ARTICLES

Dutt, A., Jain, S., Shah A. and, Thomas S. "Watching markets work: The dramatic events of 8 November 2016":
The article studies the impact of Trump presidency and demonetisation of high denomination notes on market microstructure and volatility of Indian Financial Markets [Link]

Dutt A., Sane, R. and, Thomas, S. "Direct participation in the Indian equity market: First estimates of some basic facts":
This article analyzes the household participation and portfolio choices of investors in Indian Equity markets [Link]

Dutt A., Pattanaik, A., and, Zaveri, B. "Measuring outputs v. outcomes: Did the restriction on foreign investment in Rupee-debt work?":
The article studies the statistical impact of policy changes on quality of Indian debt markets [Link]

CERTIFICATIONS

Advanced Data Science and Data Visualization - Cornell University: Relevant Coursework: Data Visualization, Predictive Data Analysis, Risk Modeling, Optimization techniques in decision making, Applied Machine Learning

Chartered Financial Analyst: CFA Level II cleared, Level III candidate

EXPERIENCE

Roc Capital **New York**
Data Science Associate *Jan, 2019–Aug, 2019*

- Identified real-estate gentrification across all US census tracts trends using a GIS driven framework and statistical learning tools
- Built and deployed a proprietary application to capture and predict human error and firm-wide risk in assessing individual mortgages, using decision trees, forests and gradient boosting to sort and intelligently select across 550 features
- Built a fuzzy string matching framework using Natural Langugage Process principles to geotag, over 140 million physical addresses to their google maps API geolocation

Thomson Reuters - Pricing Services **New York**
Quant Intern - Corporates (Emerging Markets and High Yields) *May,2018–Dec, 2019*

- Developed a cross-asset platform capable of pricing exotic sovereign and corporate debt securities, callable bonds and reverse convertible notes
- Interpolated intelligent yield curves using credit default spreads and issuer ratings for high yield sovereigns and corporates based on Nelson-Siegel, Svensson and b-spline term-structure models
- Implemented a gradient descent algorithm to model univariate, bivariate and exponential splines by minimizing least squares (The project was used for modeling term-structure of the reference rates dollarized economies such as Panama)
- Modeled the illiquidity premium for a portfolio of high-yield bullets, sinking fund and floating rate bonds on using market volatility metrics (VIX) and lagged underlying stock returns

Awarathon **Mumbai**
Data Scientist *May,2017–Jul,2017*

- Modeled smart learning modules and effectiveness gauging systems for financial literacy programs using grading networks and difference-in-differences regression

Finance Research Group, Reserve Bank of India **Mumbai**
Associate (Research) *Jun,2014–May,2017*

- Spearheaded event-study analysis of corporate actions providing research insights on investor behavior and insider trading patterns on high frequency market volumes data
- Built high-performance parallelization functions, using clustering algorithms for execution of decision trees on financial datasets analyzing portfolio allocation choices
- Provided research insights on investor behavioral patterns using K-nearest neighbour matching and stratified sampling
- Created a data warehousing and visualization platform for processing portfolio data for over 14.6 million investors spanning over 10 years and containing 236 billion records using Hive and tableau
- Implemented backtesting algorithms, to construct tick-level limit order book from high-frequency market volumes data to study the impact of fama-french parameters on the quality of stock returns
- Spearheaded key optimizations for trade upload and storage processes and wrote bulk processing APIs, which reduced intra-day position calculation time by 75% and trade validation time by 50%.

Nomura Holdings Inc **Mumbai**
Analyst *Jul,2013–Jun,2014*

- Estimated the default risk for over 1000 firms using Merton's distance-to-default KMV model and panel-data regression techniques
- Created a systemic risk index for over 300 firms using Granger Causality, Expected Shortfall and Conditional VaR metrics using high-frequency market volumes data
- Developed pipelines for logging and warehousing daily accounts payable transactions for the firm
- Created databases and downstream pipelines to maintain trade inventory for the EMEA operations of the firm as a part of the global finance architecture
- Developed and tested an, Enterprise Resource Planning module for automated trade reconcilliation

COMPUTATION SKILLS

Languages: C, C++, Python, R, SQL, VBA

Softwares/Libraries/OS: Linux(Arch/Debian/Ubuntu,Gentoo), AWS, Tensorflow, PyTorch, Theano, PyPI, Gurobi, Cvxopt, Scipy, Git, OxygenXML, Hadoop, Spark