Table of Contents

R231632 - Senior Data Scientist – Impact Assessment and Investment Optimization

Liyu Gong

205-569-4654 | wallykung@gmail.com

SUMMARY

- 5-year experience as a data scientist in dealing with both structured and unstructured data on large dataset, applying machine learning techniques to solve business problems
- Experienced in A/B testing, Time Series Forecasting, Azure Databricks, Process Automation, Productionizing Models, Customer Segmentation, Data Visualization
- Proficient in Python(sklearn/numpy/pandas/plotly/etc.), R(ggplot2/caret), SQL, SAS

EDUCATION

M.S. in Statistical Science **May 2018**
Duke University**, GPA: 3.93/4.0**

B.S. in Actuarial Science (Minor: Mathematics & Economics) **December 2015**
The Ohio State University, **GPA: 3.96/4.0**

WORK EXPERIENCE HIGHLIGHTS

Protective Life Insurance **March 2020 - present**

Senior Data Scientist

- Owned all aspects of a new payment prediction model for annuity products using survival analysis in Python, established processes for defining key metrics, hypothesis testing and implementation. The new model achieved 96% actual-to-expected ratio by capturing realistic customer behaviors and cash flows.
- Partnered with multiple cross-functional teams (engineering, BI, etc.) to build a weekly sales forecast dashboard using time series models tracking multiple lines of insurance products, achieved a 5% average error rate. Since launched, the tool has saved more than $300,000 overstaffing costs in the underwriting staff planning.
- Designed dashboards and visualization to translate the daily performance metrics and trend of the accelerated underwriting system into accessible visuals and communicated insights for both technical and non-technical end users
- Developed payment pattern prediction model for universal life insurance products in SAS and achieved 98% actual to expected ratio.

Protective Life Insurance **July 2018 – March 2020**

Data Scientist

- Researched and developed a predictive model approach to classify the mortality risk of the life insurance products using logistic regression in pyspark. Reduced the stress testing runtime by condensing the mortality tables into more accurate and efficient blocks.
- Evaluated and integrated third-party data into life insurance application risk triage model, increased the f1 score for preferred risk class by 1.6% using XGboost model.

Laboratory Corporation of America **May - August 2017**

Data Scientist Intern

- Built a claim denials classification model using gradient boosting tree in Python and achieved 89% prediction f1 score by utilizing both internal and external data.
- Implemented TF-IDF and Naïve Bayes algorithm in Python to identify patients with targeted mutations by mining unstructured lab test result notes, achieved 93% precision.
- Built a user-friendly and comprehensive business report dashboard for connecting the database and visualizing monthly lab test volume and distributions using Python PYQT4.

Sunjeev Venkateswaran

vsunjeev93@gmail.com
+1-979-422-2788

Education

Texas A&M University — College Station, TX
- *PhD, Chemical Engineering (focus on control theory);* ***GPA: 4.00*** — *Aug. 2016 – May 2021*
  Advisors: Prof. Costas Kravaris & Prof. Benjamin Wilhite

Indian Institute of Technology, Madras — Chennai, India
- *Bachelor of Technology, Chemical Engineering;* ***GPA: 8.66/10.0*** — *July. 2015*
  Minor in Management

Skills

- **Domain Expertise:** Mathematical Optimization, Control Theory, Estimation, Machine Learning
- **Programming:** Python, Java, GAMS, MATLAB, SQL, R, Calibre, Shell
- **Libraries:** Pandas, Numpy, Sqlite3, Scikit-learn, Pulp, OR-Tools, CPLEX
- **Operating Systems:** Linux, Windows
- **Version Control:** Git

Work Experience

- **Intel Corporation** — Remote
  Data Scientist/Software Research Engineer — *June 2021 - Present*
  - Developing algorithms using mathematical programming (LP, MILP) to solve large scale optimization problems in photomask design for cutting-edge technology nodes
  - Automated the entire photomask design workflow using python to maximize throughput in Intel's mask manufacturing facilities
  - Created and managed SQL databases to accurately track photomask development
  - Awarded the Department Divisional Award for increasing the throughput of photomasks by more than twofold and accelerating training of new hires

- **Texas A&M University** — College Station,TX
  Graduate Research Assistant — *Sept 2016 - May 2021*
  - Designed multiple statistical process monitoring and fault detection algorithms for stochastic nonlinear processes
  - Designed a robust observer-based fault diagnosis scheme for continuous and discrete-time deterministic nonlinear systems
  - Built algorithms to identify optimal operating points in micro-reactors using optimal control theory and mathematical programming
  - Created a framework to identify optimal catalyst distributions in heat-exchanger reactors using mathematical optimization

- **Bayer** — St. Louis,MO
  Operations Research Data Scientist Co-op — *June 2019 - Dec 2019*
  - Built a scheduling algorithm to maximize throughput in Bayer's centralized seed packaging facilities
  - Wrangled, cleaned and integrated different data sources in Python to develop streamlined inputs to the scheduler
  - The scheduler was shipped to production and reduced packaging time by about 90% during 2020 planting season.

- **Gyan Data Private Limited** — Chennai, India
  Consultant — *Dec 2015 - March 2016*
  - Developed algorithms (combination of MIP and meta-heuristics) to solve large scale capacitated vehicle routing problems

Honors & Volunteering

- Technology Development (TD) Department Divisional Award, Q4 2022
- Reviewer, IEEE Transactions in Automatic Control
- Reviewer, Chemical Engineering Research & Design
- Problem Statement Committee, INFORMS 2020 Student Competition
- Graduate Student Award, AIChE 2020- Catalysis and Reaction Engineering Division
- Texas A&M Utilities Challenge 2019 Runner Up

Select Publications

Please visit my google scholar for a more detailed list

- **S. Venkateswaran**, B.A. Wilhite, C. Kravaris. "Functional observers with linear error dynamics for discrete-time nonlinear systems." *Automatica,143, 110420*
- C. Kravaris and **S. Venkateswaran**. "Functional observers with linear error dynamics for nonlinear systems." *Systems and Control Letters,157, 105021*
- **S.Venkateswaran**, M.Z. Sheriff, B.A. Wilhite, C.Kravaris,"Design of functional observers for fault detection and isolation in non-linear systems in the presence of noises".*Journal of Process Control, 108, 68-85*
- **S.Venkateswaran**, Q.Liu, B.A. Wilhite, C.Kravaris,"Design of Linear Resdiual Generators for Fault Detection and Isolation in Non-Linear Systems". *Int. J. Control, 95 (3), 804-820*
- **S.Venkateswaran**, B.A.Wilhite, C.Kravaris,"Identifying optimal catalyst distributions in heat-exchanger reactors using optimal control theory". *I&EC Research, 59 (13), 5699-5711*
- **S.Venkateswaran**, B.A.Wilhite, C.Kravaris,"Optimal heating profiles in tubular reactors with solid-phase axial wall conduction for isothermal operation". *AIChE Journal 65 (11)*
- **S.Venkateswaran**, C.Kravaris, B.A. Wilhite,"Analysis of Solid-Phase Axial Heat Conduction upon Hot-Spot Formation in a one-dimensional Microreactor". *Chemical Engineering Journal 377, 120501*

Work Authorization

- Authorized to work in the US through H1B Visa

Hang(Hank) Zhu

Hank.zhu.hz@gmail.com https://www.linkedin.com/in/hankzhu/
(979) 422-4937 https://github.com/hankzhu4291

TECHNICAL SKILLS

Programming: SQL, Python, R, NoSQL
Tools: AWS, Sagemaker, S3, Redshift, Airflow, Jenkins, MLflow, Spark MLlib, Shell/Scripting, Linux
Machine Learning: Classification, Regression, Clustering, Dimensionality Reduction, NLP, Recommender System
Data Visualization: Tableau, Seaborn, ggplot2, Plotly, D3.js, Google Data Studio

PROFESSIONAL EXPERIENCE

Data Scientist II, Amazon, New York, NY Sept. 2022 - Current

- Collaborated with Product, Engineer team to build machine learning driven Ads targeting strategy
- Communicated with business stakeholders to translate ambiguous problems into science problems
- Designed efficient ETL pipelines to process billion level data from different sources using SQL and Spark
- Built innovative dimension reduction models and scalable and reproducible segmentation models in Sagemaker
- Delivered model solutions with concise data visualization to drive Ads activation and creative insights

Data Scientist II, Simulmedia, New York, NY March 2019 – Sept. 2022

- Collaborated with Product, Engineering team to build media Ads measurement/prediction/optimization platform
- Communicated with internal stakeholders to translate product needs into data science problems
- Built efficient ETL pipeline in SQL and built scalable and reproducible machine learning models in Python
- Designed statistical experiments and monitoring system to validate model effectiveness in post-production
- Delivered analytical solutions with effective data visualizations to enable data-driven decision making

Data Science Consultant, DIA Associates, New York, NY Sept. 2018 – March 2019

- Provided data-driven strategy consulting for the Digital Marketing department in American Express
- Wrangled, explored and summarized billion-level website clickstream data using Hive
- Derived insights using Excel and Python to resolve problems and recommend digital marketing strategies
- Built and optimized visitor segmentation model and designed experiments to validate models
- Delivered analyses and strategies by creating data visualizations and dashboard to help decision-making

Data Scientist Intern, Singleplatform, New York, NY June 2018 – Sept. 2018

- Collaborated with sales and marketing teams in understanding and analyzing business requirements
- Wrangled and integrated 1.5 million rows of data from SQL, Yelp-API, BigQuery, and Salesforce
- Developed machine learning models on AWS to predict customer conversion rate in order to improve sales
- Engineered features and tuned hyperparameters to increase performance metric AUPRC from 0.54 to 0.79
- Created data visualizations and derived insights to help management make better business decisions

Data Scientist Intern, Simulmedia, New York, NY May 2017 – August 2017

- Researched and derived insights by wrangling 4 million rows of data using SQL
- Predicted household ethnicity to improve the precision of TV Ads performance measurement
- Tuned machine learning models to increase classification performance metric AUPRC from 0.65 to 0.8
- Delivered presentations with data visualizations to drive better business decision-making
- Deployed code to production for business use to allow customers to reach target audiences more accurately

EDUCATION

Master's Degree - Industrial Engineering (Statistics track), Texas A&M University May 2018
Bachelor's Degree - Industrial Engineering, Huazhong University of Science and Technology May 2016
CFA Level I passed

Rutvi P Bhatt

8, Browning Road Arlington, MA 02476 · 781-290-9604 · rutvib@bu.edu

Education:

Boston University College of Engineering:

MEng Biomedical Engineering (Dec ’19) | **B.S. Biomedical Engineering (**Concentration in Technology Innovation) | May ‘19

Awards: Dean Lutchen Research Fellowship, Cum Laude, Dean’s List

Skillset: Project Implementation, Strategy design for scaling, Analytical & Problem Solving, Leadership and Management, Effective Communication and Storytelling

Work Experience/Internship:

Accenture LLP, Boston, MA — Mar’20 – Present

Consultant in Life Sciences Practice, Full-Time

- Designing and implementing data driven solutions for large multinational companies to provide their clients with a seamless omnichannel experience and customized marketing.
- Ideating and co-designing work practices and platforms, that will enable different functional groups within a company such as IT, Brand/Marketing, Sales, Insights & Analytics groups to work together.
- Implementing AI/ML models and decisioning engines to build solutions that allow for sales and marketing channels to work effectively to provide the right message, using the right channel at the right time for client’s customers.
- Defining strategy for scaling a program from local to regional to global level by performing capability analysis (People, Technology, Data, Processes) and determining cost to implement across multiple countries and brands.
- Championing new programs and initiatives by building governance and change management models to help with adoption within the company for launches.
- Advising on KPIs, organizational design and funding structure for launching programs and projects for clients.
- Managing multi-vendor ecosystems and projects to allow for structured, cross functional collaboration and implementation.

Medtronic, Danvers, MA — May’19 – Dec’19

Advanced Manufacturing Engineer, Co-op

- Designing, developing, and optimizing manufacturing processes for multiple components of a heart valve delivery system.
- Improving manufacturability of delivery catheters by reducing time and costs of production while error proofing development steps.
- Manufacturing and prototyping delivery components using injection molding, braiding and fusing for discrete builds.
- Performing DOEs, mechanical testing (Tensile, 3 point bending, Young’s Modulus) & statistical analysis for characterization of components.
- Performing comparative studies amongst 3D printed ceramic molds, acrylic molds and steel molds in order to determine feasibility of molding.
- Designing molds in SolidWorks and finding acceptable tolerances on molds to produce parts within specifications.

Partnership for Global Health Technologies, Zanzibar, Tanzania/ Boston, USA — Jan’18 – May’ 19

Research/Design Engineer, Research

- Applied image processing and machine learning techniques to create a non-destructive, affordable, point of care device for assessing quality of mRDTs (malaria rapid diagnostic tests) and to confirm validity of its results.
- Collected and processed microscopy samples and dry blood samples (DBS) to compare against mRDT results for enhancing machine learning.
- Extracted DNA from DBS samples and performed PCR in order to examine parasite presence, density, and species.
- Drafted reports on findings of quality control guidelines for malaria diagnostics and worked with malaria elimination programs in Zanzibar.

Lyndra Inc., Watertown, MA — May’17 – Aug’17

Platform Development Engineer, R&D Internship

- Formulated ultra-long dosage, polymeric tablets for oral medication at bench-scale and large-scale using twin-screw extruder.
- Manufactured and optimized drug dosage forms using Infrared Welding technology.
- Implemented multiple testing methods to evaluate mechanical and chemical properties of materials and led workshops to discuss collected data. Testing methods included: *Instron Tensile Testing, 4-Pt. Point Bending, water absorption/mass loss, thermal studies with DSC/TGA.*
- Drafted SOPs and data analysis templates to prepare documentation for future use.
- Performed large volume data analysis using Excel and Minitab software.

Harvard Medical School in collaboration with BWH and MGH, Boston, MA — May’16 – Sep’16

Design Intern, Internship

- Fabricated a low-cost, single use sugar testing lancet for diabetics to prevent accidental pricks or reuse of needles in low income countries.
- Designed lancet structures in CAD using solid works and printed them with 3D extrusion printer in order to explore different geometries.
- Studied filament materials like ABS/PLA properties to choose appropriate material for building structures.
- Prepared structures with reverse molding methods and evaluated mechanical properties such as sharpness of the fabricated lancet.

RASHMI SINGH

508-723-6563 |singhrashmiajay@gmail.com|New York| www.linkedin.com/in/rashmi-singh-ajay

WORK EXPERIENCE

Lead Data Analyst, Citibank N.A, United States **Jan 2022-Present**

- Performed quasi-experiments to asses impact of Zelle,Wallets usage on customer engagement and revenue.
- Designed and implemented A/B experiments using Mastercard's Test & Learn tool to support the product teams by analyzing the impact of new launches on key product metrics
- Leveraged Random Forest models in python to identify significant features that are predictive of customer engagement.Used these features to train propensity to click model for recommending non-marketing activites
- Built ad-hoc summary reports on large scale cards and retail customers data (~ 32 million) using advanced SQL

Manager, Data Science, Bristol Myers Squibb, United States **Jun 2020-Jan 2022**

- Developed a text clustering model using k-means to identify major reasons for extensions of quality incidents.
- Built a document search tool to identify similar quality incidents using RWMD similarity NLP model
- Collaborated with Manufacturing and Supply Chain teams to reduce carrying cost for finish goods by 17% by leveraging gradient boosting machine learning model for predicting the true lead time of a batch.

Data Science Co-Op, Bristol Myers Squibb, United States **Jan 2020 – May 2020**

- Developed statistical model using Spotfire to perform process capability analysis for improving process yield.
- Created ETL pipelines using Python to standardize data from different sources like AWS Redshift, Oracle and MySQL

Data Science Intern at Akamai Technologies, United States **May 2019 – Aug 2019**

- Designed a Causal Inference framework using Matching method to estimate effectiveness of marketing campaign
- Developed a propensity score model using logistic regression in R to be used for Matching & IPTW algorithms.

Data Analyst at Larsen & Toubro, India **Jul 2016 – Jul 2018**

Fraud Detection in Employee Travel Claims

- Developed a fraudulent travel claims detection model using Random Forest in Python with f1 score of 0.87
- Generated five-point summary statistics within each segment to identify outliers & set threshold for future cases.

Bank Data Analysis

- Analyzed 1.5 million internal and 3rd party bank transactions to identify anomalous activities using advanced SQL.
- Built interactive dashboards for the same using Microsoft BI tools like MS SQL, SSAS, SSIS and Power BI**.**

EDUCATION

- **Worcester Polytechnic Institute (WPI), Worcester, MA** **May 2020**
  Master of Science in **Data Science** **GPA 3.72/4**
- **University of Mumbai** **June 2016**
  Bachelor of Engineering in Information Technology **CGPA 8.66/10**

ACADEMIC PROJECTS

Progressive Approximate Answers for Aggregation Queries **Sep 2019 – Dec 2019**

- Optimized response of aggregation queries on a huge dataset of 1 million rows using Spark Scala.
- Performed progressive sampling of the data to report quick approximate results until the results converged.

Topic Modeling using Natural Language Processing (UTC Connecticut) **Sep 2019 – Dec 2019**

- Designed a topic modeling system using Latent Dirichlet Allocation (LDA) to identify issues from unstructured chats.
- Performed tokenization, lemmatization POS tagging on chats and designed features using Bag of Words & TF-IDF

TECHNICAL SKILLS

Programming Languages: Python (NumPy,Scikit-learn,PyTorch, Pandas,Keras, Gensim, spaCy, Matplotlib) R, C#, Java
Databases: MS SQL, PostgreSQL,Oracle,Amazon Redshift,MySQL
Visualization tools: Tableau, Power BI, QlikSense, TIBCO Spotfire, Google Datastudio, Python Dash
Big Data Technologies: Apache Hadoop,Hive, Spark, Cassandra, Elasticsearch,MongoDB

CERTIFICATES & ACHIEVEMENTS

- Tableau Data Scientist Badge | Hackerank Python Certificate |
- A Crash Course in Causality: Inferring Causal Effects from Observational Data by University of Pennsylvania

SRAVAN ROY AKKINENI

sravanroy247@gmail.com | https://github.com/sravanroy | https://www.linkedin.com/in/sravan-roy/ | 470-399-4968

SUMMARY

Data Scientist with experience in developing scalable solutions for various verticals by analyzing data and delivering ML models and dashboards. Developed recommendation engines and data pipelines for Fortune 50 clients.

EDUCATION

Georgia State University, USA

Master of Science – Data Science & Analytics (*GPA : 4/4*)

NIT Jamshedpur, India

Bachelor of Technology – Electronics & Communication

EXPERIENCE

Walmart *Data Scientist, Remote, USA* Sep 2021 - Present

Tech stack - Python, GCP, Airflow, BigQuery, NLKT, Spacy, Git, DVC, Streamlit, Element ML

- Created ETL pipelines in Python for Walmart brands and scheduled automated jobs in Airflow to increase the efficiency of campaign management by 25%
- Developed category dashboards in Streamlit and deployed to google app engine to provide real time insights to stakeholders
- Increased the lift by 35% by improving accuracy of ensemble of ML classifiers that drive recommendations for chatbot using causal inference techniques
- Reduced the ML model retraining and deployment time by 30% by implementing DVC for all production ML models
- Improved the load time of daily reports by deploying cron jobs to automate data reduction in the sftp servers using Python

Verizon *Data Analyst, Florida, USA* Sep 2020 - Sep 2021

Tech stack - Python, GCP, Advanced SQL, Azure ML, Tableau, Oracle SQL, Pulsar Client

- Reduced the costs by 30% by implementing a dynamic route planner for telecommunication client engineers using ML
- Decreased the % of inefficient trips made by field engineers to 40% by analyzing the vehicle info, gmaps api and their past history
- Predicted the fleet travel route for engineers with 72 % accuracy based on their previous tasks and travel route patterns
- Reduced the Call Drop Rate (CDR) for customers by analyzing the 4G and 5G call parameters and identifying the loopholes in the network parameters that caused consistent call drops

Mu Sigma *Data Scientist, Bangalore, India* May 2017 - July 2019

Tech stack - Python, SQL, RedShift, Flask, AWS, R, Tableau, ML, R Shiny, Kubernetes

- Generated a consistency module in Python to detect anomalies and flag them based on historical trends
- Implemented a demand forecasting tool in R by analyzing and modeling historic data using ARIMA
- Improved sales reporting speed by 30% by transforming static reports to automated dashboards with live SQL queries

TECHNICAL SKILLS

Programming:	Python, R, SQL
Databases :	SQL Server, MySQL, NoSQL, Hadoop, Redshift
Tools :	Git, Docker, Flask, DVC, Tensorflow, PyTorch, Spark
Cloud :	AWS (S3, Sage Maker, Athena, Glue, EC2, Lambda, EMR,), Azure (Azure SQL, ADLS, ADF)
Visualization :	Tableau, Power BI, Excel, Streamlit
Core Competencies :	Supervised and Unsupervised ML, Regression, A/B Testing, Optimization, Customer Segmentation, Hierarchical Modelling, Inferential Statistics, Churn Analysis, Market Basket Analysis, Causal Analysis

PROJECTS *(project links are embedded)*

- Designed a recommendation system using **Word2Vec** to recommend similar products based on purchase history
- Generated a **Tableau** dashboard to analyze the music trends using Spotify's official API to scrape and collect songs data
- Designed an **R Shiny** dashboard to predict the rating of board games with Linear, Ridge, Lasso, PCR, and PLS methods
- Developed a flask web app to predict output from a LGBM model and hosted on **Azure** by spinning up Docker container
- Built an **LSTM** model to predict the sentiment of a given text and deployed the flask app on Heroku Cloud service

Farzad Daneshgar, PhD

(817) 471-0979 | farzad.daneshgar.ofcl@gmail.com

EDUCATION

University of Maryland, College Park, MD PhD,
Operations Research, May 2018

Dissertation: A Comprehensive Mixed-Integer Programming Model to Optimize the Performance of Freeway Service Patrol Programs

University of Maryland, College Park, MD
Master of Science, Applied Mathematics & Scientific Computation, May 2018

Minor: Management Science

University of Texas at Arlington, TX
Master of Science, Transportation Engineering, December 2011

Sharif University of Technology, Tehran, Iran
Bachelor of Science, Civil Engineering, June 2010

PROFESSIONAL EXPERIENCE

Senior Research Scientist (Oct 2022-March 2023) **| Research Scientist II** (Sept 2020-Sept 2022)**, Amazon**, Customer Service Worldwide Capacity Planning**,** Seattle, WA (Sept 2020-March 2023)

- *Optimize Customer Service Network Contact Allocation Considering Cost and Risk:* Evaluated risk and cost elements, quantified the metrics and variables, established the strategy, and developed the optimization model for distribution of contacts inside the network and network expansion which minimizes for the cost and risk while satisfies all security/operational/capacity requirements.
- *A Methodology to Estimate the Thresholds of Overtime for Customer Service Capacity Planning:* Developed the methodology to predict the number of attainable overtime slots to be posted for upcoming scheduling optimization; determined the limit for the slots by analyzing historical data on posted and filled overtime slots and training a parametric model to calculate fill rates while handling censored data by applying a flexible constrained machine learning model.
- *Downstream Impact of Customer Service Experience:* Designed the methodology and developed the statistical framework to quantify the impact of customer service experience metrics such as waiting time and abandonment on customers by applying causal inference techniques and estimating the average treatment effects.
- Provided technical support for long-term headcount planning and short-term workforce scheduling.

Sr. Operations Research & Adv. Analytics Specialist, BNSF Railway, Fort Worth, TX (July 2018- Sept 2020)

- *Prediction of Trains Estimated Time of Arrival:* Participated to develop XGBoost models to predict train ETA at destination hubs, performed data analysis and feature engineering, and predicted possible delays due to train parking and maintenance events. Also, formulated an optimization model to imitate the plan for trains movement in the network to estimate delay incurred by lower priority trains.
- *Service Design:* Developed a mixed-integer programming model and heuristic algorithm to improve the railroad blocking design given new traffic volume and capacity constraints.
- *Locomotive Demand Prediction:* Developed an ARIMA time series model to forecast locomotive demand.
- *Locomotive Fuel Consumption:* Developed a multivariate regression model to estimate and compare fuel efficiency of locomotive models equipped with different technologies and saved several million dollars in fuel consumption and purchasing the correct technology.
- *Set out Damaged Railcars to Mechanical Shops:* Processed and analyzed the network data, developed the algorithm and the mathematical formulation to determine the optimal set out by minimizing the deviation from the original set out while satisfying shop capacity and business constraints.
- *Grain Empty Shuttle Swap:* Participated to develop an optimization model to determine candidate empty shuttles for destination swap by estimating potential run time savings.

Graduate Researcher, University of Maryland, College Park, MD (Jan 2016-July 2018)

- *Analysis and Optimization of CHART (Coordinated Highway Action Response Team) Operations:* Prepared the proposal, processed and analyzed network data, developed the mathematical model and heuristic algorithm to design routes for emergency incident response units, determine optimal fleet size and fleet assignments into routes, minimize total incident response duration, minimize operation cost, and estimate traffic delay savings.
- *Optimization of Emergency Traffic Patrols Operations:* Prepared the proposal, extend the mathematical model to determine the most efficient patrol coverage area (given an underlying transportation network) to avoid unnecessary patrolling, determine the routing (considering multiple coverage) for patrolling area and determine

non-patrolling area by considering elements such as incident frequency, operation cost, average patrolling and dispatch response times.

GIS Research Specialist, Center for Advanced Transportation Technology, College Park, MD (Feb 2015- Jan 2016)

- Developed a GIS algorithm to calculate nationwide volumes for INRIX TMC-Based network using HPMS (Highway Performance Monitoring System) AADT datasets.
- Created a model and GIS tool to project networks in public sector legacy linear referencing systems onto networks with private sector spatial segmentations.
- Guidance on Qualifying Benefits of Traffic Incident Management Strategies: Performed B/C analysis of traffic incident management strategies based on set of standard MOE's applying various methodologies.

Consultant, Maryland State Highway Administration, Baltimore, MD (Oct 2014-Sept 2017)

- Implemented Geospatial analysis and crash data analysis.
- Executed traffic modeling and simulation tasks.
- Directed transportation planning, traffic engineering operations, safety and Before and After studies.
- Conducted field observations, traffic and travel time (GPS) data collection.

Data Analyst Intern, Federal Highway Administration, Research Center, Mclean, VA (Sept 2012-Feb 2014)

- *FHWA Highway Safety Information Systems Project*: Evaluated incident data, performed statistical analysis on incident database using SAS, updated HSIS guidebooks, and provided technical support.

Graduate Research Assistant, University of Maryland, College Park, MD (Jan 2012-Aug 2012, Feb 2014-Jan 2015)

- *I-95 Corridor Coalition: Vehicle Probe Project Validation of INRIX Data*: Prepared data collection plans and revised sensors segment planning based on the actual implementation, participated in deployment of Bluetooth monitoring systems to collect ground-truth travel time data, prepared INRIX and sensors collected data for further process applying large database tools, prepared reports for deployments using evaluation tools.

Graduate Research Assistant, University of Texas at Arlington, TX (Mar 2011-Sep 2011)

- *Development of a Transportation Sustainability Index for Urban Communities, North Central Texas Council of Governments (NCTCOG)*: Developed a set of MOE's for sustainable transport, quantified MOE's using multiple choice tests, and performed a case study in city of Plano, Texas

PUBLICATIONS & CONFERENCE PROCEEDINGS

- Farzad Daneshgar, Stephen P. Mattingly, and Ali Haghani, "Evaluating Beat Structure and Truck Allocation for The Tarrant County Courtesy Patrol" Transportation Research Record, Network Modeling, 2013, Volume 2, pp 40-49
- Farzad Daneshgar, Sahar Nabaee, "Dynamic Message Sign Location Configuration Using Linear Model" Proceeding of Transportation Research Forum, 2013
- Shahram Bohluli-Zanjani, Farzad Daneshgar, and Siamak Ardekani, "Development and Validation of a Direct Mode Choice Model" Transportation Planning and Technology, Volume 37, Issue 7, 2014
- Farzad Daneshgar, Ali Haghani, "Joint Mixed-Integer Model to Minimize Incident Response Time in Freeway Service Patrol Program", Transportation Research Board Annual Meeting Compendium of Papers, 2016, Paper No. 16-4269
- Farzad Daneshgar**,** Kaveh Farokhi, and Ali Haghani, "A Conflation Methodology for Two GIS Roadway Networks and its Application in Performance Measurements" Transportation Research Record (2018): 0361198118793000
- Hyoseuk Chang, Ali Haghani, and Farzad Daneshgar, "A Proposed Generalized Framework for The Identification of Contributing Factors Which Affect Hazardous Traffic Conditions on Freeways" Transportation Research Board Annual Meeting Compendium of Papers, 2018, Paper No. 18-05601

COMPUTER SKILLS

- Python, SQL, MATLAB, SAS, R, FICO Xpress, CPLEX, Gurobi, ArcGIS, Palantir, Tableau, Power BI, Azure, JavaScript, HTML, Linux, LaTeX, Microsoft Office

ADDITIONAL INFORMATION

- U.S. Citizen
- Qualification Exams, University of Maryland: Operations Research, Probability, Management Science

Chendong Li

Phone: *(403)-667-1364*, *Email: lichendonger@gmail.com*
Address: 569 St Giles Rd, West Vancouver, Canada

SUMMARY

I have over 7 years of experience working on data science and advanced analytics. Currently, I am a data science team lead in Cisco Chief Data & Analytics Office developing and productize enterprise AI/ML solutions to enable sales growth and guide sales process.

Competencies: machine learning, NLP, Python/R, deep learning, Keras/TensorFlow, PySpark, SQL, NoSQL, Docker, GCP

WORK HISTORY

As a part of Enterprise AI organization transforming Cisco's go-to-market, specifically sales teams and partner channels, I lead the end-to-end development of data science projects including problem formulation, model prototyping, development and deployment: engage with stakeholders to frame business requirements into data science problem, develop automated pipelines for data pre-processing and feature engineering, research state of art industry and academic solution, select most promising model candidates, define benchmarks and evaluation metrics, train and evaluate the model, finetune the model using hyper parameter optimization, and finally deploy the model using the microservice architecture. I leverage the power of Python, Spark along with various ML tools such as Docker, DataProc, Cloud Run, App Engine, Vertex AI and Airflow to produce highly scalable and easily deployable real-time predictive solutions integrated into Cisco sales and operation.

Data Science Team Lead, Enterprise AI at Cisco Systems, Vancouver — March 2022 – Present

- Led the development of Cisco recuring revenue forecasting for global sales team. Leveraging the hierarchical categorical embedding using deep learning, the forecasting model is deployed and adopted by Cisco Finance to set up the half-year goal target of Cisco 17K sales representatives, which accounts for approximate 20% of Cisco revenue
- Engage with stakeholders to frame vague business problems into scientific problems. Lead the team to develop prototyping solutions

Lead Data Scientist, Enterprise AI at Cisco Systems, Vancouver — Oct 2021 – March 2022

- Led NPS response propensity project for Cisco quarterly NPS (Net Promoter's Sore) surveys. Built and evangelize the propensity score model so that the model solution was successfully gained business buy-in and operationalized into the customer feedback workflow.

  With this model solution, Cisco increased the number of respondents by 41% and increased reachability by 11%, as well as reducing customer fatigue by reducing our invite pool by 10%.
- Mentor and coach junior data scientist and intern

Senior Data Scientist, Enterprise AI at Cisco Systems, Vancouver — Jun 2021 – Oct 2021

- Developed opportunity score model to predict which deals are most likely to win by current quarter and deployed the model in Cisco Salesforce Platform
- Developed causal analysis and run experiment to rigorously estimate the dollar impact of guided sales tools for Cisco sales team

Data Scientist, Enterprise AI at Cisco Systems, Vancouver — Jun 2019 – Jun 2021

- Developed Siamese network model with character-based LSTM to identify duplicate customer name and address records for 24 million+ records of Cisco Customer Registry
- Built synthetic labelled data using customer names and develop Bi-LSTM+FastText model for customer name entity recognition

- Achieve 10x more duplicate throughput at only 50% cost; Presented the project in Data Symposium (Cisco internal, widely attended annual event)
- Prototype data science solution using ML classifier that solves Cisco's lingering problem of identifying if hardware devices are still active or decommissioned in the customers' install base

Data Scientist, Co-Founder, QuestFlex Inc, Vancouver **May 2018 – Jun 2019**

QuestFlex is a technology start-up that utilizes artificial intelligence including NLP, big data, machine learning and deep learning to empower conventional business. As one of the founding members, my primary responsibility is leading the project of cryptocurrency asset management and developing machine learning based high-frequency trading system to generate profitable risk-adjusted return in the cryptocurrency market

- Lead the entire product life-cycle including research, design, data processing, feature engineering, model selection, back-testing, implementation, and performance evaluation
- Build machine learning models from scratch to production (including random forest, isolation forest, GBM, feedforward neural net, and LSTM with RNN) using Python to detect market anomalies and predict market patterns
- Tools & Technologies Stack: Python, sckit-learn, Keras, TensorFlow, NoSQL, Node JS, AWS cloud service

Various Analytical Roles for Energy Market, Calgary **May 2014 – Feb 2018**

1. Analyst, Strategic Planning & Forecasting, ATCO — Aug 2016 – Feb 2018
2. Analyst, Simulation and Forecasting**,** Alberta Electric System Operator (AESO) — Jan 2016 – Aug 2016
3. Trading Analyst**,** Electricity and Natural Gas, TransAlta — May 2014 – Oct 2015

- Provided quantitative and fundamental forecasting for electricity & natural gas market to support proprietary energy trading, corporate budget, business development, and strategic planning
- Developed statistical and ML models to predict demand and market price in electricity & natural gas market
- Extracted market insights and present findings & strategies to enable data-driven decision makings
- Tools & Technologies Stack: R, MATLAB, PYTHON, SQL

SKILLS

Machine Learning: Classification and Regression (Tree-based models, Logistic Regression, Naive Bayes, Nearest Neighbor Algorithms). Unsupervised learning (Hierarchical and K-means Clustering)

Deep Learning: Convolutional NN, Recurrent NN, Hierarchical Attentional Networks, Transformers

Statistical Methods: Point and Interval estimation, Hypothesis Testing, Analysis of Variance and Regression, Time Series Analysis, Bayesian Inference, Causal Inference

Software and Programming Languages: Python (Scikit-earn, Xgboost, Lightgbm, Keras, Tensorflow, PyTorch, Scipy, Numpy, Pandas, Matplotlib), PySpark, Dash, R, Matlab, SQL (PostgreSQL, MySQL), NoSQL (MongoDB), Docker, Git

GCP: DataProc, Compute Composer, Airflow, Cloud Run, App Engine, Dataflow, BigQuery

AWS: Sagemaker, S3, EC2, Redshift

EDUCATION

M.A (Thesis-Based), Economics **Sep 2011 – Sep 2013**

University of Calgary

B.S, Physics (Major) **Sep 2007 – Jul 2011**

B.S, Finance (Minor) **Sep 2008 – Jul 2011**

East China Normal University, Shanghai, China **Dec 2013 – May 2014**

CFA Level III Candidate, CFA Institute **Jun 2015 – Jun 2019**

Jia Liu

24 Passaic Ave, Hasbrouck Heights, NJ 07064 | 484-929-3063 | liujia1991.work@gmail.com

SUMMARY

5+ years experienced analytical consultant with strong technical skills in the pharmaceutical industry. Advanced practical and educational focus on SAS/R/SQL programming, data analysis, data mining, and pharmaceutical consulting. Expertise in learning and using new technologies very fast. Process highly developed time arrangement, multi-project scheduling, and teamwork skills.

EDUCATION

Lehigh University, Bethlehem, PA, U.S. **08.2014-05.2016**

Master of Engineering in Industrial & Systems Engineering

Zhejiang University of Technology (ZJUT), Hangzhou, Zhejiang, China **09.2010-06.2014**

Bachelor of Engineering in Mechanical Engineering, minor in Financial Management

SKILLS

- Certifications: SAS Certified Base Programmer for SAS 9, SAS Certified Advanced Programmer for SAS 9
- Proficient in SAS, SAS Enterprise Miner, SQL, R, Microsoft Office
- Experienced in Python, Tableau, PowerBI

RELATED EXPERIENCE

Senior Associate Data Scientist, Definitive Health - Analytical Wizards, Iselin, NJ, U.S. **02.2019-now**

- Have supported multiple clients, including Sanofi, Regeneron, Novartis, Otsuka, Eisai and G1, etc.
- Hands-on experience in different kinds of projects, including Segmentation Analysis, Treatment Pathway, Referral Dynamics, Market Mix modeling, ROI analysis, Trail Accelerator, Compliance Optimizer, etc.
- Familiar with most of pharmaceutical related data, including sales data, prescription data, promotion activity data, specialty pharmacy data, APLD data, Symphony claims data and EMR/EHR data
- Proficient in advanced analytical methodologies in SAS, R and SQL to meet customer needs

Key Projects:

- Set up segmentation lists on account & physician level by using K-means clustering and Latent class clustering, based on market size, market share, HCP capacity, promotion activities, etc.
- Hands on multiple EDA analysis including demographic and comorbidities analysis on patients level, site of care and referral pathway analysis on prescriber level, and payer channel and access on payer level
- Hands on experiences in statistical modeling (e.g., regression model, time series model) for multiple promotion activities, optimizing different promotion channels, creating marketing mix models to maximize marketing ROI
- Worked on real world medical claims data to conduct patient level analysis, such as design patients' persistency predictive model, and other deep-dive data analysis to help client make strategic business decision which could improve patients' engagement, retention, and growth
- Worked on Flatiron EMR/EHR data to analyze advanced NSCLC treatment pattern by PD-L1 biomarker to help client have a better understanding of current market performance of drugs and drug classes used to treat NSCLC patients; Identify key market leaders within each PD-L1 expression level segments and uncover potential opportunity for Novartis launch brand to compete effectively; Profile NSCLC patients based on age, stage, gender and PD-L1 status
- Worked on APLD data to predict patient-specific discontinuations before occurrence; Create alerts for Specialty Pharmacies Support Teams to intervene and prevent discontinuations; Improve patient duration of therapy
- Built business insights dashboard to visualize, track, and monitor sales trend across geography, identify any of potential business change
- Support advanced analytics team on multiple ad-hoc business requests by programming in SAS/R with large and complex datasets (e.g., Rx sales, promotion, demographic, alignment etc.)

Data Analysis Consultant, KMK Consulting Inc, Morristown, NJ, U.S. **07.2017-01.2019**

Key Project: Novartis Pharmaceuticals Corporation, East Hanover, NJ, U.S.

- Worked as a data analytics consultant at Novartis Data analytics and strategy team, had major responsibility of developing standardized data integration process and data validation process across multiple data vendors
- Automated processing of delivering weekly/monthly/goaling transactions files, worked closely with IC (Incentive compensation) team to set up goals and payout model by cycle

Inventory Analyst, Kelly Services, Center Valley, PA, U.S. **12.2016-06.2017**

- Worked at client site (Olympus America) to perform data analysis/validate through SAP and Salesforce on their inventory account with physical field inventory audit results from a third-party company

December 30, 2022

To Whom It May Concern,

I am writing this letter of recommendation on behalf of Shyam Menon, Data Scientist at Hallmark.

Having known Shyam for almost a year as his manager, I can attest to his data science and business acumen.

During his time working with me, he has impressed me with his ease to coach, willingness to take on big projects and excellent presentation skills. He is proactive, delivers projects on time and calm under extreme pressure. He stays current in programming languages and machine learning techniques. And as a fully remote team, I most appreciate his ability to stay linked with me and the team, work independently, and lead complicated projects. Shyam's positive attitude, analytical and statistical skill set, and professionalism makes him a joy to know and a true asset to the team.

He has developed the company's home-grown solution to measure AB lift analysis replacing a tool the company contracted annually. He also has helped the data science and engineering team adopt new technologies including AWS modeling and storage solutions.

I am confidently able to recommend Shyam Menon based on this experience. Also, I am more than happy to provide further details about Shyam if needed. I can be contacted directly at www.linkedin.com/in/tammy-green-ds1234.

Sincerely,

Tammy Green

Tammy Green
Sr. Data Scientist II

Shyam Menon

12409 Bunche Road, Fairfax, VA 22030 | (703) 656-3620
menonshyam@hotmail.com | Linkedin.com/in/menonshyam703

Education

University of Colorado – Boulder May 2021 - Present

- Master of Science in Data Science *(expected graduation, May 2023)*

Ross University School of Medicine August 2019 – July 2020

- Doctor of Medicine candidate *(left due to COVID-19 pandemic change to online learning)*

Virginia Commonwealth University

- Bachelor of Science in Science

Certifications and Specializations

Microsoft Corporation

- Microsoft Certified: Azure Data Fundamentals

Johns Hopkins University

- The Data Scientist's Toolbox - R Programming

University of California – Davis

- SQL for Data Science

University of Michigan

- Python 3 Specialization
- Statistics with Python Specialization

Relevant Skills

- Proficient at Python, R Programming, SQL, Tableau, Power BI, Java, Git, and Microsoft Office
- Proficient with various cloud-based AWS products (SageMaker, Simple Storage Service [S3], Athena)
- Experienced in using NumPy, Pandas, Scikit-Learn, SciPy, Seaborn, Matplotlib, Geopandas, Contextily, Mapclassify, OpenCV, Tidyverse, ggplot2, R markdown, and Jupyter notebooks

Work Experience

Hallmark Cards, Inc. March 2022 – Present

- Data Scientist, Insights & Analytics Manager
  - Deliver key benefits by implementing on-time, on-budget, and on-target solutions through cross functional partnerships across the business practice
  - End-to-end development and management of Performance Insights Tool (PIT), an applied statistics and experimentation (e.g. A/B testing) platform:
    - Combination of Ball Tree Nearest Neighbor matching process and Linear Mixed Models to account for random effects in lift analysis
    - Cost-benefit, price elasticity, and risk analysis to provide holistic business insights that drive impactful change
  - Built clustering and segmentation models to reengineer supply chain and staffing efforts, which saved 10MM in spend
  - Built machine learning models to predict consumer behavior on multiple dimensions of spending, preference, and segmentation (45MM consumer base)
  - Built longitudinal and trend long/short range machine-learning and regression forecast models
  - Manage and train a team of Data Scientists and DS interns on a variety of projects

        - Assign team members projects, troubleshoot any inaccuracies/anomalies, provide suggestions/review of statistical models, and retool/refresh statistical models
    - Administer continued recruitment processes for the Data Science organization
        - Triage, interview, hire, and on-board new team members
    - Management and implementation of multiple Microsoft tools (Azure DevOps, Forms, Power Automate, SharePoint, Teams, etc.) for the Data Science organization

Smithsonian Institution April 2019 – August 2019

- Technical Specialist
    - Strategy and reengineering of technical programs. Managed 30 volunteers and implemented efficiency changes that increased guest retention and interaction rate.

National Institutes of Health (NHLBI) May 2013 – August 2013

- Research Intern
    - Analysis of genomic and health-related large data sets. Produced statistical models for virology (adeno-associated viruses) and gene therapy research.

The Pennsylvania State University August 2010 – May 2012

- Research Data Scientist
    - Experimental design and performance analytics (A/B Testing). Leveraged advanced statistical techniques to produce targeted biofuel technology. Implemented process improvement initiatives and root causes analyses.
    - Publication: Callus induction and RAPD analysis of Simarouba glauca DC. African Journal of Biotechnology, 13(53), 4766-4774. https://www.ajol.info/index.php/ajb/article/view/126034

United States Department of Agriculture December 2011 – January 2012

- Research Intern
    - Nutrition and health data analysis. Produced statistical models for targeted protein analysis and phytonutrients enhancement of tomato fruit.

Leadership Experience and Activities

- Smithsonian Institution - Udvar Hazy Air and Space Museum 2008 – 2018
    - Managed and directed volunteers to conduct and construct a learning environment for guests
- Young Speakers Club of Greater Washington 2006 – Present
    - Mentor to younger members and practice public speaking
- South Asian Student Association (SASA) Executive Board 2012 – 2015
- President of Future Business Leaders of America (FBLA) 2008 – 2010
- FIRST Robotics 2008 – 2010
    - Won the Automotive and Innovation Award
    - Built robots for nationwide FIRST Robotics competitions

Volunteer Experience

- Library of Virginia June – August 2020
    - Volunteered in the “Making History” Transcribe Project – helped transcribe historical documents into digital text
- Capital Area Food Bank January 2020
    - Managed and helped organize food stocks for distribution to local community programs

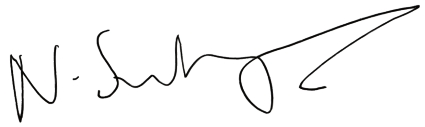
Microsoft Certified
Azure Data Fundamentals

SHYAM MENON

Has successfully completed the requirements to be recognized as a Microsoft Certified: Azure Data Fundamentals.

Date of achievement: December 30, 2021

Microsoft



Satya Nadella
Chief Executive Officer

Certification number: I096-3897

Harshit Manektalia

(213) 551-3367 https://www.linkedin.com/in/harshitmanek/ h160983@usc.edu

EDUCATION

UNIVERSITY OF SOUTHERN CALIFORNIA Los Angeles, CA
Master of Science in Computer Science (Artificial Intelligence) ***GPA: 3.95*** Aug 2021-May 2023
- Coursework: Algorithms, Artificial Intelligence, Machine Learning, NLP, Big Data Mining, Deep Learning

UNIVERSITY OF DELHI, CLUSTER INNOVATION CENTRE Delhi, INDIA
Bachelor of Technology in IT and Mathematical Innovations ***GPA: 83.15%*** ***Rank: 3*** Aug 2013-May 2017
- Coursework: Probability & Statistics, Python, C++, R, MATLAB, SQL

EXPERIENCE

GENENTECH – *NLP powered Recommendation Engine* South SF, CA
Data Scientist Co-op May 2022–Present
- Spearheading R&D of an NLP based Recommendation Engine to facilitate clinical trial submissions to regulatory authorities.
- The engine uses fine-tuned state-of-the-art Large Language Model and reduces the submission time by ~16 weeks per trial.

ELUCIDATA – *Drug discovery using Artificial Intelligence* Delhi, INDIA
Data Scientist Feb 2019–Jul 2021
- Led 6-member team to successfully develop world's largest curated multi-omics data lake consisting of over 1.5 million datasets, spanning more than 5000 diseases and 50 major public data repositories along with an intuitive analytical interface in R shiny.
- Facilitated labs at UCLA, Yale, Pfizer, and Flagship Pioneering in their research with biomarker identification, drug repurposing, etc. by applying ML and Statistical Models to multi-omics datasets and Computer Vision techniques to imaging datasets.
- Developed a Recommendation Engine to suggest relevant Literature & Datasets by fine-tuning state-of-the-art NLP techniques (*BERT, huggingface transformers*) on bioRxiv and PubMed corpus and using NER tagging for tissue, gene, disease, and drug.

INDIQUS TECHNOLOGIES – *Cloud Resource Forecasting and Capacity Planning* Delhi, INDIA
Data Scientist Mar 2017–Feb 2019
- Designed and Implemented VM Capacity Planning using Multivariate Time Series Forecasting, powering intelligent recommendations to scale cloud infrastructure, resulting in 20% monthly cost reduction for customers and reduced churn rate.
- Implemented sales forecasting, customer segmentation, and market basket analysis techniques to assist the sales team, resulting in a 35% month-on-month increase in revenue through new lead generation and successful cross-selling and up-selling strategies.

DEFENSE RESEARCH & DEVELOPMENT ORG., *SAG LAB – R&D Org. under Ministry of Defense* Delhi, INDIA
Machine Learning Research Intern Jun 2014–Dec 2015
- Developed tools for Signature Verification and Face Detection using various ML and DL techniques (*SVM, LDA, CNNs*).
- Designed and presented a neural network-based hybrid classifier at ICEMS 2016, Jaipur, India.
- Implemented, Carried out Attacks, and Ameliorated cryptographic schemes such as Cayley Purser, Wang and Hu, and SVD.

ACADEMIC PROJECTS

Celebrity Recognition in Video using Semi-Supervised Learning
- Designed and Implemented an ensemble of two semi-supervised algorithms by combining graph-based similarity measures of MAD with the Transductive Support Vector Machine of LDS.
- Trained and Tested the ensemble on Cropped Yale Database and IMFDB (Indian Movie Face Database), obtaining accuracies of 85% and 65% respectively, thus, outperforming both MAD and LDS when considered individually.

Computer vs Human GO using Artificial Intelligence
- Built a Computer versus Human GO game using Alpha-Beta Pruning, Monte Carlo Tree Search, and Deep Learning.

TECHNICAL SKILLS

- **Languages**: Python (*Keras, TensorFlow, scikit-learn, PyTorch, transformers, PySpark*), R, C/C++, MATLAB, Unix, JS, SQL
- **Tools:** Git, Markdown, Jupyter Notebook, LATEX, AWS, GCP, Docker, Hadoop, Spark, MLflow

AWARDS & GRANTS

- Awarded Pilot Grant of £10000 by The Alan Turing Institute and The University of Exeter for research titled "An Integrated Data Science Pathway for Informed Drug Discovery in Motor Neurone Disease".
- Won International NEUROHACK 2022, a 4-day competitive hackathon promoting brain health.
- Ranked 3rd in class (Bachelor's degree, University of Delhi).

Akshat Sanghvi

Champaign, IL, USA 61820 | akshat72525@gmail.com | 217-721-8155 | LinkedIn | GitHub

PROFESSIONAL EXPERIENCE

Walmart Global Tech | Data Scientist | Intern May 2022 – Aug 2022

- Developed machine learning pipeline to forecast Cases per Trailer (CPT), potentially saving **$10M** annually through optimized inventory management and saving **4320 man-hours per year** at the Distribution Center level
- Partnered with Data Analytics team to extract and aggregate **90,000+** data points from GCP BigQuery using SQL, performed data cleaning and preprocessing to build a CPT forecast model with **94%** accuracy for non-seasonal periods using R/Python
- Exceeded target accuracy by **13%** for seasonal periods using XGBoost by utilizing KMeans/Ward Hierarchical algorithms to cluster **400+** seasonal items and used their demand as exogenous variables in the model
- Empowered business partners to make accurate decisions by developing a Tableau dashboard of forecast results

University of Illinois Urbana-Champaign | Data Scientist | Research Assistant Jul 2022 – Dec 2022

- Aided Prof. Aravinda Garimella in her efforts to address issues caused by unethical use of Artificial Intelligence Systems
- Conducted data analysis and feature selection on Experian credit data for over **1M+ consumers** containing **500+ attributes**, in order to identify key predictors of credit default
- Built an ensemble classification model using Python to predict credit default, with a ROC-AUC score of **0.65**

Supercut Engineering Works | Data Analyst | Co-op May 2019 – Aug 2020

- Worked with clients and cross-functional teams to define and understand the functional requirements of the product
- Designed Tableau dashboards showcasing KPIs, allowing inventory managers to make data-driven stocking decisions
- Utilized KNN and fuzzy matching techniques for targeted marketing by comparing external customer data sources with the company's customer database, resulting in a **10%** increase in customer acquisition rate and **$300k** additional revenue in 2019
- Led a team of two data analyst interns to develop predictive maintenance models to reduce downtime and increase machine availability by **15%**, resulting in an annual savings of **$100k** in maintenance costs

University of Illinois Urbana-Champaign | Graduate Teaching Assistant Jan 2022 – Present

- Guided **18 students** per semester in their Senior Design project by offering project management and technical support
- Collaborated with professors to design an innovative course curriculum and engaging assignments for **200+ students**

TECHNICAL SKILLS

- **Programming:** Python, R, C/C++, Java, MATLAB, SAS, SQL, NoSQL, Neo4j, VBA
- **Python Libraries:** NumPy, Pandas, SciPy, PyTorch, TensorFlow, OpenCV, XGBoost, Gurobi, SciKit-Learn, Keras
- **ML & DL Techniques:** Linear Regression, Logistic Regression, Random Forest, SVM, Neural Networks, CNN, GAN
- **Analytics & Others:** Tableau, PowerBI, ETL, Hypothesis Testing, A/B Testing, Spark, Hadoop, Docker, Kubernetes
- **Cloud technologies:** Google Cloud Platform (BigQuery), AWS (SageMaker, S3, QuickSight, Athena)

ACADEMIC PROJECTS

Land Cover and Crop Type Segmentation [GitHub] Nov 2022 – Dec 2022

- Created pixel-level labels based on crop-type maps from Cropland Layer images provided by USDA
- Incorporated UNet to segment Corn, Soybeans from other crops on RapidEye Satellite image with **85%** Pixel accuracy

PASCAL YOLO Object Detection [GitHub] Mar 2022 – May 2022

- Implemented custom YOLO loss function on pre-trained ResNet model to detect objects with mAP of **0.38**
- Applied score-thresholding metric to draw boxes around and removed overlapping boxes using non-max suppression

Spotify – AI in Business [GitHub] Sep 2021 – Nov 2021

- Developed a content-based recommender system using KNN to suggest similar **five music tracks** to users
- Performed feature importance using the Random Forest on over **160k tracks** containing **60+ features**
- Predicted skip probability of a user with an **88%** accuracy by hyperparameter tuning of LightGBM using Optuna

EDUCATION

University of Illinois Urbana-Champaign, Champaign, IL Aug 2021 – Apr 2023

Master of Science in Industrial Engineering, Concentration – Data Science & Machine Learning *GPA – 3.6/4*

Coursework – *Deep Learning for Computer Vision, Statistical Learning, Cloud Computing, Network Analysis*

University of Mumbai, Mumbai, India Aug 2017 – Jul 2021

Bachelor of Engineering in Engineering *GPA – 3.89/4*

Alec J. Soudry

(330) 388-6830 | alecsoudry@gmail.com | Cincinnati, OH

EXPERIENCE

North American Bancard October 2020 – Present

Senior Data Scientist - Enterprise Risk Management Troy, MI

- Led *Data Science & Reporting* in designing an automated approval system to reduce friction amongst potential/onboarding customers, while concurrently automating the underwriting team's responsibilities.
  - Decreased the amount of time credit analysts spent manually reviewing merchant applications by ~54%.
- Performed analysis and guided research to build a high-ticket buffer system for new merchants that decreased unnecessary alerts to the underwriting team by 19%.
- Developed interactive Tableau dashboards to visually enhance analytic findings for managing directors to holistically monitor 5 million+ data points and provide insight for KPI trends.
- Established an end-to-end analytics spectrum that prescribes actionable insights to leadership with visibility into how business drivers were trending against goals, and alignment to the business strategy.

FIS Global June 2018 – October 2020

Data Scientist – Advisory & Marketing Cincinnati, OH

- Developed and operationalized predictive models using consumer segmentation data that generated prioritized credit-union members exhibiting the highest propensity to buy.
- Designed A/B and multivariate testing methodologies for evaluating impact and measuring performance of digital campaigns, observing an increase in marketing responses by 26%.
- Redefined consumer segmentation to improve selection and servicing metrics, resulting in an uplift in targeting by 40% and enhanced customer experience metrics such as NPS score, retention rate, etc.
- Created marketing mix, marketing response, multi-touch attribution, and forecasting models to identify sales growth by customer, institution, incentive, etc. to identify sales trends and client spending behavior.
- Implemented reporting on multi-variant success metrics by market, business segment, website and campaign level and developed strategic initiatives to optimize GTM marketing campaigns.

Synchrony Financial May 2017 – June 2018

Data Analyst - Amazon Stamford, CT

- Constructed tests and performed analysis to determine the ideally formulated campaign, as well as the subsequent promotional deals on store/gift cards, special financing arrangements, & Spend-and-Get offers.
- Worked with the credit and client development teams to identify and maximize customer likelihood of upgrading to unsecured or PLCC, resulting in a conversion rate increase of 34%.
- Built customer attrition models that improved monthly retention by 11% for customers who were likely to churn by offering relevant products and personalized offerings.
- Used customer segment analysis to develop marketing strategies based on path of entry and user behavior within the website and the Amazon card program.
- Increased analytical rigor and gained process efficiency across the *Measurement and Analytics* team by reducing team SLAs through process development/improvement, and automating reports via Alteryx & Google Analytics - decreasing reporting time by 5-10% per week
- Formed Power BI dashboards for senior management by pulling consumer and business reports using optimized SQL queries to track MoM, YoY and YTD changes using key financial profitability metrics.

EDUCATION

University of Cincinnati Cincinnati, Ohio

Bachelor of Science, Business Analytics & Computer Science

- GPA: 3.8/4.0, National Merit Scholar, 8x Dean's List Recipient
- Relevant Coursework: Algorithm Design, Application Development, Data Mining, Data Visualization, Database Design, Decision Models, Forecasting, Statistical Modeling, Systems Analysis & Architecture

SKILLS

Languages: Python, SQL, R, Java, Scala, Ruby, JavaScript, C#, C++, TypeScript (Most to Least Proficient)

Platforms: **Apache** (Cassandra, Hadoop, Hive, MapReduce, Spark), **AWS** (Athena, DynamoDB, EC2, EMR, Glue, Lambda, QuickSight, Redshift, SageMaker, S3), **Google** (Analytics, AutoML, BigQuery, GCP), **Microsoft** (Azure (Databricks, Synapse), Excel, Power BI, SSMS/SSRS)

Databases & Tools: Docker, Git, H2O, Jira, Kubeflow, Kubernetes, Looker, MariaDB, MLflow, mongoDB, MySQL, Neo4j, Oracle, Postgres, Qlik, SAS, Sigma, Snowflake, Tableau

Sreevatsan Agoramurthy

(602) 475-5155 | sreevatsan.agoramurthy@gmail.com | linkedin.com/in/sreevatsan-agoramurthy/
New York, NY, USA

Education

Arizona State University - WP Carey School of Business *Tempe, Arizona, USA*
Master of Science in Business Analytics, *GPA: 3.9*

Anna University - College of Engineering, Guindy *Chennai, Tamil Nadu, India*
Bachelor of Engineering in Computer Science and Engineering

Technical Knowledge

Languages	Python, C++, C, C#, Java, R, PHP, SQL, T-SQL, NoSQL, JavaScript
Tools	SAS, Hadoop, Apache Hive, Tableau, Power BI, Snowflake, Airflow, Azure ML
Skills	Machine Learning, Clustering, Regression, SVM, Deep Learning, Business Intelligence, Data Visualization, Statistical Analysis, Data Warehousing, Big Data Analytics, A/B Testing, Data Engineering

Experience

WeWork
New York, New York, USA

Senior Data Scientist – Full Time **March 2022– Present**

- Performed A/B tests on membership purchases & user journeys using SQL & Python leading to a **12% increase** in checkout conversion rate.
- Defined **driver & success** metrics for the member facing product group and aligned strategy with company KPIs.
- Designed, developed and deployed multiple **experiments & multivariate** tests to drive pricing & promotional strategies.
- Deployed a comprehensive **CLTV** model for members using NLP and regression models that led to a **24% increase** in subscription renewal rates.
- Quantified the dollar value of member acquisition through **exploratory** data analysis, **clustering** user behavior and analysis of **session** data.
- Mentored a team of **3 product analysts** by delegating tasks, organizing offsites, and brainstorming best practices in product and statistical analysis.

Data Scientist - Full Time **June 2021 - March 2022**

- Constructed **data pipelines** by performing ETL & ELT on source data, defining business logic and implementing reporting tables to analyze **user behavior** on the e-commerce platform
- Constructed data models with data engineers to analyze **user engagement & stickiness** on member facing products.
- Shaped **product strategy** for member facing product by defining **user personas** through quantitative analysis, statistical modeling and clustering on user engagement data.
- Minimized **cart abandonment** rates by 28% through the design, deployment and monitoring of multiple **experiments**.
- Owned and instrumented the end-to-end event **tracking framework** (using Segment) for user interactions on the e-commerce platform, member facing product groups and internal tools

DoorDash
Tempe, Arizona, USA

Data Scientist - Contract **October 2020– May 2021**

- Accelerated the performance of data pipelines by 40% using **parallelizable & scalable API calls** with Python, JSON & MS SQL from 3 sources – JIRA, Workday and Genetec (Security Software).
- Developed a **BI solution** using MS SQL Server and Tableau dashboards to proactively identify risks, evaluate platform safety and support risk mitigation strategies.
- Developed **back-end system** for an in-house web application which facilitated event scheduling and seating arrangements in line with seating constraints due to COVID-19 regulations.

Dell Technologies
Tempe, Arizona, USA

Data Scientist - Internship **May 2020– October 2020**

- Eliminated design related issues early in the product development cycle by 10% by **identifying anomalies** in motherboard component failures using **neural networks** through TensorFlow and Domino.
- Slashed warranty related costs to the company by 8% through the development of a complex **predictive model** using machine learning and statistics to predict component failure rates in products.

Moumita Hanra
Phoenix, Arizona, 85013
Phone: 623-999-8818 ,
925-385-8231
moumita.hanra@gmail.com
www.linkedin.com/in/moumita-h-7009737

H1B transfer needed

Professional Experience Key Highlights:

- Data science professional with Master's degree in Quantitative Analysis and 10 plus years experience in the field of Advanced Analytics and Machine Learning (ML); Excellent interpersonal and communication skills with expertise coding in **Python**, **SAS**, **SQL**, **R**, **SPSS**, **Excel**, **OCTAVE**, **VBA, Jira, Azure cloud**
- Data acquisition, data mining, merging, data cleaning, data visualization, data transformation, feature selection and data engineering of both Structured and unstructured data
- Supervised ML algorithms like **MMM**, **Linear, Logistic, Polynomial regression, L1 (Lasso) and L2 (Ridge) regularization techniques, Boosted trees, AdaBoost, XGBoost, CatBoost, LightGBM, Random Forest, Support Vector Machines(SVM), Neural networks, KNN, Naive Bayes**, Time series models like **ARIMA, AR, MA** models, **Structural Equation Modeling**, from scratch ML algorithms like **Gradient descen**t and **Normal equation regression**.
- Unsupervised ML techniques like **K-means, Hierarchical, Hybrid of K-means and Hierarchical, Latent class** and **SVM** based clustering, Dimensionality reduction techniques using **Principal component analysis, Multi-dimensional scaling, KNN,** and **LDA.**
- Natural Language Processing (**NLP**) for Sentiment analysis and Text analytics**, CHAID, CART, Recommender Systems,** Optimization using **VBA, ANOVA, ANCOVA, MANOVA, Brand Equity Scores, Correspondence Maps, Max Diff Scaling, TURF, Fair Value, Risk Reward, Lead Lag analysis, Anomaly detection, Conjoint,** and **Van Westerndorf pricing.**
- Experience in industry verticals – Healthcare, Pharmaceutical, CPG, Education, Food Service, Video Games, Manufacturing, Retail, Insurance, Financial Services, Telecommunications, Automotive, others.

Education Summary:

- **Master of Science** in **Quantitative Analysis** from **University Of Cincinnati**, June 2009
- Bachelor of Engineering in Instrumentation, University of Mumbai, Aug, 2005

Technical Skills:

- **Python** (Advanced knowledge):

    - **Python libraries** include **Pandas**, **Numpy**, **Matplotlib**, **Plotly**, **Seaborn**, **Scikit learn**, **SciPy**, **NLTK**, **BeautifulSoup**, psycopg2
- **SAS** (Expert knowledge)
- **SQL**(Expert), **PostgreSQL**(Advanced)
- **SPSS, R, Excel** (Advanced), **VBA**
- **Powerpoint** (Advanced), **OCTAVE, MATLAB, arcGIS**
- Amazon **Redshift** database
- **Tableau**, **Power BI**

Visa Information:

- H1B transfer needed
- Canadian PR: can begin work immediately remotely till H1 approved

Work Experience:

Senior Manager Data Science, Transunion, Phoenix, Arizona, Jan 2022 to March 2023:

- Conducting Market mix modeling for retail and financial services industries
- MMM using Bayesian methods and frequentist methods using multiplicative log log model; prior values are calculated using prior business knowledge of the variable, using ground truth studies or referring to coefficients of similar variables across industries or specific to industry
- Conduct MMM to predict top driver of sales and other KPIs that are also potential drivers of sales like online traffic and Google query volume
- The potential drivers are ad metrics like TV, out of home, online video, online display, social media Pinterest, Facebook, Instagram and radio etc as well as econometric variables
- Use Python for EDA, and custom software with R as backend for conducting MMM and deploying results and forecasts.

Manager Data Science, Blend 360, Phoenix, Arizona
June 2021 to Dec 2021:

- Query data via SSMS SQL in Microsoft Azure db and do data engineering, and conduct market segmentation of physicians or patients to form their meaningful groups based on demographics, medical claims data, Sales RX data, SDOH and others using cluster analysis.

- Analyze data using qualitative and advanced analytics techniques to target physicians via proper MCM and other sales tactics by pharma company sales reps.

- Conduct market mix modeling (MMM) to predict and optimize RX scripts of doctors to increase sales of drugs and to optimize sales tactics around TV, digital etc as well as physician detailing by sales reps.

- Conduct XGBOOST and other classification models to target doctors who are highly likely to write RX or engage with digital campaigns.

Data Science Manager, Genpact, Phoenix, Arizona
Feb 2021 to June 2021:

- Worked for GSK in helping them target the right physicians who are likely to prescribe RX and identifying gap and sweet spots using classification algorithms like Logistic Regression and XGBoost.

- Segmentation of physicians using K-means and Hierarchical cluster analysis in Python

Statistical Programmer, Arizona State University, Phoenix, Arizona
Aug 2018 to Sep 2020:

- Working in the Health Research department and serving mainly government insurance payers; Working with Healthcare claims data to pre process data and solve key client objectives like reducing payer costs, coded in SAS macros to automate many processes.

- Conducted **Segmentation** to create distinct groups of Physicians or patients for purposes of targeting, used **Random Forest, Boosted Trees, Logistic Regression** and other ML to detect high risk patients using multiple predictors, worked on **Linear regression** models to predict drivers of Payer costs with predictors like demographics, insurance, claims record, diagnosis, test results, pre-existing conditions and treatment history including medications.

- Weighted **Social Network analysis** of physician networks to show the most influential physicians, and create optimal shortest specific paths from one physician to another to be followed for optimal cost allocation between different centers; worked on **arcGIS** for visualizing and representing geospatial data; worked on **ANOVA, MANOVA, t-tests, chi-square tests** to compare patient or doctor groups.

Advanced Analytics Consultant, McClelland Engineers, Mumbai, India
April 2015 to June 2018:

- Incorporating supervised ML techniques to assess the impact of marketing measures and locations on Sales of manufacturing machines and making recommendations on how and where to target potential clients for best sales.

- Allocation of factory workers in the correct stage of production and calculate the best number of workers at certain times of the day for incurring least production costs.

- Conduct statistical tests to compare performance of factories in different locations

Associate Director Advanced Analytics, Symphony Health, Willow Grove, PA
Oct 2013 to Feb 2015:

- Served as a lead operations research manager leveraging primary and secondary data for the pharmaceutical industry and supervised two on-shore and two off-shore team members.

- MMM, Detail and Message Effectiveness using regression modeling, Physician Segmentation using Cluster analysis and Adherence studies using Survival modeling and classification methods like logistic regression

- Detailed study of Oncology patients at various stages of progression and analytics to derive cohorts as well as calculate Persistence/Adherence using survival modeling

Manager Advanced Analytics, Starcom Mediavest, San Francisco, CA
Feb 2012 to May 2013:

- Conducted powerful Market Mix Modeling (MMM) and ARIMA modes to predict sales using potential predictors like TV GRPS, Digital Display, Digital video, Google Search, Word of Mouth, social media data, pricing promotions, Campaign, customer attitudinal data, and socio economic data and demonstrated Diminishing return curves of paid media and earned media and develop ROI numbers for each media type.

- Design VBA based cost allocation Optimizers that determine best budget allocation strategy among TV and digital and create weekly contribution charts and forecasted sales.

- Performed advanced analytics on Cookie level data and Digital media data to demonstrate the drivers of end metrics like product purchase/ product trial.

- Interacted with data vendors to discuss data requirements, addressing data questions and proposing data layout suggestions.

JERRY PUNNOOSE

PROFESSIONAL SUMMARY

Tenacious and sincere professional with 4 years of experience in Statistical Analysis, Process Optimization, and Data Science to deliver innovative technical solutions for clients in the Life Sciences industry. Committed to leading a visionary team in managing data mining and analysis projects to drive actionable insights for business growth.

WORK HISTORY

Project Lead | Decision Science, 04/2021 to Current
Axtria - Ingenious Insights - Berkeley Heights, NJ

- Leading teams focused on client-specific models providing insights on current campaign gauging and providing recommendations on optimizing spend and targeting.
- Formulated statistical models of problems, relating constants and variables, conflicting objectives, and related numerical parameters to optimize annual marketing budget and improve YoY topline revenue by $11M.
- Onsite consultant supporting Omnichannel Reporting team at a major pharmaceutical company in creating field performance dashboards and automated data models for validation scripts for these dashboards, thus eliminating process inefficiencies by 90%.
- Worked on a Segmentation & Targeting project to optimize the sales force efforts for a Vaccines portfolio leading to optimal allocation of FTE resources.

Senior Marketing Analyst, 11/2018 to 03/2021
Outcome Health - New York, NY

- Directly interfaced with clients from Pharma industry and media agencies to align on campaign measurement methodology and provide insights on campaign effectiveness and brand performance.
- Conducted post-program management of pharmaceutical campaigns by developing a codebase to evaluate campaign impact measurement, identifying participants generating high ROI of 6:1-10:1 and over $7M in incremental revenue.
- Collaborated with Sales & Product Marketing team members to define campaign objectives (e.g. ROI vs Awareness) by conducting quantitative analyses on IQVIA datasets and creating weekly KPI decks to monitor performance across areas of interest and growth.
- Directed A/B Testing to measure patients' behavior in doctor's offices and increase ad awareness among patients by providing qualitative surveys, concluding that ad-aware patients were 12% more likely to buy the brand.

Predictive Analytics Intern, 09/2017 to 05/2018
Delaware North

- Utilized Unsupervised Learning algorithms including Clustering and

CONTACT

Address: Secaucus, NJ 07094
Phone: 716-275-4561
Email: jerry.shaji@gmail.com

PROFILE

- https://www.linkedin.com/in/jerry-shaji-punnoose/

EDUCATION

Master of Science, Industrial Engineering | Operations Research, 06/2018
University At Buffalo, SUNY - Buffalo, NY

Bachelor of Technology, Mechanical Engineering, 05/2016
Vellore Institute of Technology - Vellore, India, India

SKILLS

- Marketing Mix Modeling
- Predictive Modeling
- Market Access Analytics
- Pharma Data - HCP, APLD, Claims etc.
- SQL
- Python
- Data Science
- Segmentation and Targeting
- Promotion Response Modeling
- Experimental Design & Hypothesis Testing
- Program Management
- Digital Analytics

CERTIFICATIONS

- Python Programming | General Assembly, September 2021
- SQL and Database Design | Udemy, March 2019
- Tableau 10 Essential Training | LinkedIn, December 2017
- SAS Certified Base Programmer | SAS Institute, October 2017

Association Rules to develop 5-10 Product Combos leading to a sales revenue increase of $200,000 during the baseball season.

- Developed a Linear Programming optimization model to solve labor scheduling conflicts at a restaurant, optimizing the working hours and shifts of employees that translated to $150,000 in savings annually.
- Created a dynamic dashboard for tracking real-time resort performance of capturing the bookings and availability of vacation resorts by using SAS VA and writing the SQL queries which created multiple static excel reports delivered from the resorts.

Rachel Freeman

Durham, NC | 703-597-3419 | refreeman15@gmail.com
www.linkedin.com/in/freeman-rachel/

Data professional with 4+ years of experience in data manipulation and cleaning, data pipeline design, algorithm design and implementation, predictive modeling, and statistical analysis. Passionate about optimizing workflows and processes through development of tools or use of statistical and predictive models. Problem solver that enjoys working on collaborative teams to help organizations make data-driven decisions.

Skills

Big Data Analytics, Statistical Analysis and Modeling, Data Engineering, Data Visualization, Predictive Modeling, Machine Learning, Data Cleaning and Preparation, Communication of Data Insights

Programming Languages and Tools:

Python | R | C | Linux | GIT | SQL | PySpark | Apache Airflow | Apache Beam | Apache Pig | Agile Methodology

Google Cloud Platform (BigQuery, Composer, Dataflow, Dataplex)

Experience

Data Engineering Consultant, Accenture, July 2022 – Present

- Create DAG (Directed Acyclic Graph) Workflows in Airflow to add functionalities such as data cataloging, new data ingestion, changing data requirements, and table modifications to financial data pipeline.
- Perform ad-hoc validations and changes in tables in BigQuery to assist in data exploration.
- Discover and fix code issues in client's Python scripts, related to lack of standardization and hard coding, making production environment less susceptible to failure.
- Learned role on the fly in a startup environment on a team with many competing priorities and stakeholders.

Data Scientist, Department of Defense, July 2019 – June 2022

- Built predictive model for data analysis process requiring manual review, enabling prioritization of information.
- Formulated statistical experiment to provide product assurance before deployment to production environment.
- Modeled employee attrition using survival analysis models, specifically Kaplan-Meier curves, to determine features that affected attrition in significant ways.
- Applied clustering algorithms to unlabeled text strings after formulating word embeddings, which resulted in ability to group strings by structural similarity and ultimately assisted technical analysts.
- Cleaned and combined data from various sources because raw data was not standardized and required encoding.

Computer Science Intern Program, Department of Defense, May 2018 – August 2018

- Improved algorithm that found relevant information to a search term from corpus of text through implementation of new method using graph theory and ranking algorithms.

Education

MS in Applied and Computational Mathematics

Johns Hopkins University, January 2020 – December 2021

BS in Mathematics

Virginia Tech, August 2015 – May 2019

TABASSUM FAZEL

fazeltabassum@gmail.com• +1 612 987 9914 • https://www.linkedin.com/in/fazeltabassum/

Experienced Data Scientist in delivering **data-driven insights** and **actionable solutions**. Proficient in employing complex machine learning/ statistical techniques and adept at the communication of innovative ideas enabling collaborative thinking. Currently looking for opportunities where in I can use my scholastic background, industry experience and zeal to learn.

SKILLS: Exploratory & Statistical Data Analytics, Predictive Modeling, Hypothesis Testing, A/B Testing, Market Basket Analysis, Clustering, Visualization, Machine Learning, Python, R, SQL, Snowflake, Tableau, PySpark, Hive, Airflow, AWS

EXPERIENCE

OPORTUN INDIA, Chennai, India
Senior Data Scientist *Sep 2022 - Present*

- *Model Monitoring:* Developed model monitoring report on Jupyter for ongoing risk evaluation model. Designed model assessment framework for validity and stability by including metrics – **AUROC, PSI, CSI, KS and lift charts**
- *Model Development:* Developing ML model for assessing risk of Direct Mail prospects **~5MM** monthly on **PySpark ML**

DISCOVER FINANCIAL SERVICES, Chicago, IL
Senior Data Science Analyst *Sep 2020 – July 2021*

Project: New Model Score Based Risk Segmentation

- Updated the two dimensional risk segmentation matrix for Direct Mail prospects with new model scores
- Optimized the risk matrix to ensure similar volume mix and KRIs with the old strategy to minimize impact on portfolio
- Revised strategy resulted in an additional **3.4MM** prospects with more stable pre and post mailing customer risk profiles

Project: Response Channel Model Prediction

- Profiling direct mail marketing channel prospects for **feature engineering** for 5+ MM customers per mailed week
- Developing a model to predict the response channel for the Direct Mailed customers using Scikit-Learn

CARLSON ANALYTICS LAB, Minneapolis, MN
Analytics Student Consultant *June 2019 – May 2020*

Client: Leading Addiction Treatment and Advocacy Organization

- Conducted **market basket analysis** to reveal co-occurring substances abused and treatments assigned to 250K patients
- Conducted **propensity score matching** in R for identifying patients similar to patients who relapsed on drugs or alcohol
- Developed **Random Forest classifier** in Python for patient's risk evaluation during on-boarding for effective treatment

Client: Mall of America

- Consolidated various data sources to create holistic view of MOA security call information **250 MM** rows using Python
- Identified patterns by **K-prototype clustering** method (Python) for well-informed patrolling routes and staff allocation

KABBAGE (FinTech Industry), Bengaluru, India
Senior Decision Analyst *June 2016 - May 2019*

Project: Underwriting/ Qualification Model Development and Strategy Design

- Consolidated data from multiple sources for 100K customers leveraging MS SQL for model development and validation
- Developed **logistic regression L1** penalty model for risk assessment of the customer using Python-Scikit Learn package
- Estimated an overall increase in the credit approvals by 10% and loan originations/year by **$20MM**
- Devised initial A/B testing framework of modified underwriting methodology integrating new data source and L1 model

Project: Europe & UK Partner Portfolio Management

- Managed end to end risk portfolio and established continual relationship with various UK and Europe banking partners
- **Optimized** incoming customer funnel using MS Excel and Python resulting in additional 8% customers in the portfolio
- Devised account management strategies utilizing **association principles** (**Apriori Algorithm**) for risk management

Project: A/B Testing Sampling Tool

- Devised a **sampling tool** to generate similar samples for A/B tests on current portfolio for statistically significant results
- Created an **interactive GUI** to facilitate conducting campaigns/ tests across teams with minimal statistical background

ACADEMIC PROJECTS

- Evaluated **'LIME' package** of Python on Titanic data - Kaggle by building a logistic regression for survival prediction
- Developed a forecasting model for restaurant customers, data from Kaggle by building a **Light Gradient Boost** model
- Implemented a framework utilizing varied **AWS Services** on incoming streaming data for real time fraud detection

EDUCATION

UNIVERSITY OF MINNESOTA, Minneapolis, MN
Carlson School of Management
Master of Science in Business Analytics *May 2020*

INDIAN INSTITUTE OF TECHNOLOGY MADRAS, Chennai, India
Dual Degree: **Bachelor's & Master of Technology** in Mechanical Engineering *May 2016*

Brian Livian

NYC

516-359-4165 brianlivian@gmail.com linkedin.com/in/brianlivian/ github.com/brianlivian

Skills & Interests

Programming: Python (Pandas, Matplotlib, Sklearn, XGBoost, Pytorch, Ray, Optuna, etc.), SQL, R, HTML
Applied Mathematics & Statistics: Topological Data Analysis, Hypothesis Testing, Optimization
Big Data: AWS (S3, Sagemaker, EC2, EMR Studio), PySpark, Azure, Databricks
Software: Tableau, Docker, Git, Excel, PowerPoint
Interests: Licensed motorcyclist, EMT, rock climber, paddle-board enthusiast

Experience

Universal McCann — New York, NY

Senior Data Scientist — *Apr. 2022 – Present*
Data Scientist — *Sep. 2021 – Apr. 2022*

- Developed machine learning models to predict and optimize marketing ROI/ROAS
- Designed a novel outlier detection method, significantly improving model results while minimizing data removal
- Designed a multiprocessing hyperparameter tuning algorithm which resulted in at least 5x improvement in tuning times
- Utilized cloud computing for computationally demanding tasks (hyper-parameter tuning and partial dependence plots)
- Wrote Python functions to streamline/ automate historically manual BI tasks
- Created ETL pipelines to automate data ingestion, standardization, and reporting
- Created data visualizations (Python & Tableau) to showcase key performance metrics and data findings
- Prepared presentations and reported model results to senior management and internal/ external stakeholders

All American Healthcare — Newark, NJ

Machine Learning Engineer Intern — *Jul. 2021 – Sep. 2021*

- AAHC is a healthcare staffing technology company with over 200,000 users and 50% YOY growth for 4 years
- Built machine learning recommendation models to provide personalized suggestions for app users
- Exported and deployed models into production via API using Docker and AWS containers
- Wrote SQL queries in PostgreSQL to acquire data via company's databases with millions of data points

S&P Global — New York, NY

Data Scientist Intern — *May 2021 – Jul. 2021*

- Web Scraped 5,000+ company financial reports and used regular expressions to extract key insights
- Used natural language processing to extract text data from company losses reports and perform statistical analysis
- Collaborated with colleagues using git, creating a seamless project workflow

Projects & Research

Traveling Salesman Project

- Wrote a script which takes a list of cities/ landmarks as input and optimizes the most efficient route between them (while taking into account the curvature of the Earth), plotting the geospatial data with plotly

Topological Data Analysis of Financial Time Series

- Demonstrated the application of TDA to predicting stock market crashes using a multivariate financial time series

Financial Impact of Tropical Cyclones on U.S. Real Estate Sector

- Quantified a relationship between the increasing frequency of hurricanes and the financial impact on Real Estate Corps.

Twitter Stock Analysis

- Webcraped 10,000+ Elon Musk tweets over an 11-year period and used natural language processing for keywords of Dogecoin and Tesla to show 44% correlation between Musk's tweets and significant change in asset price

Education

Yeshiva University, Katz School of Science and Health

M.A. Mathematics- 3.9

- Thesis: Topological Data Analysis of Periodic Time Series with Noise

City University of New York, Queens College

B.A. Mathematics- 3.6

CHENGBO DU

Jersey City, New Jersey 07302
+1 (310) 721-1702 duchengbo94@gmail.com

EDUCATION

University of California, Los Angeles Los Angeles, CA
Master of Financial Engineering, 3.9/4.0 GPA Sep 2017 - Dec 2018

- *Volunteering*: Social planning officer of the class

University of Science and Technology of China Anhui, China
Bachelor of Science in Physics, 3.7/4.3 GPA Sep 2012 - Jun 2016

- *Honors*: Outstanding Graduate in Province (top2%), Excellent Student Scholarship Gold Award (top5%)

SKILLS

- *Programming and software*: Python (Pandas, NumPy, PyTorch, PySpark, Scikit-learn, SciPy, Matplotlib), R (ggplot2, stats, dplyr, data.table), C++, SQL, SAS, Hive, AWS (EMR, EC2, Glue, Lambda, S3, Step Functions, SageMaker), Power BI, Microsoft Excel
- *Technical skills*: Machine Learning, Neural Networks, Time Series Forecast, A/B Testing, Causal Inference, Pipeline Automation

PROFESSIONAL EXPERIENCE

Amazon Seattle, WA
Data Scientist Nov 2022 – Present

- Science lead for ML-based prediction settlement model that provided propensity score/segmentation for marketing team's event-based campaigns and loyalty campaigns to drive incremental ordered purchase sales and digital engagement on Amazon homepage
- Designed and maintained ML pipeline for data integration, feature engineering, job monitoring, model training, performance evaluation, model selection, scheduled prediction and deliverables integration with AWS
- Implemented AutoEncoder and TF-IDF to deal with high sparsity and high dimensionality of customer data in trillions
- Proposed local-to-global hybrid Pangea model framework for category/product granular level prediction

DIA Associates (Onsite consultant at American Express) New York, NY
Associate, Data Scientist Aug 2019 – Nov 2022

- Collaborated with marketing team's leadership to transform marketing performance analytics into leading solutions and conduct statistical measurement analysis on major email and tele channel campaigns with annual charge volume lift over $3 billion
- Implemented supervised learning algorithms and built statistical models like Random Forest, XGBoost to generate insightful analysis on high-value products and form model-based customer targeting prioritization strategy
- Designed A/B testing experiment standard procedure and impact attribution protocol for multiple marketing channels
- Developed a new time series prediction methodology by adding COVID adjustment factors to existing time series models like Facebook Prophet, Seasonal Theta to address the structural break during the pandemic
- Integrated, transformed and visualized sophisticated user data and then automated the metrics dashboard to keep tracking of multiple model-based trigger campaigns' weekly performance

OmniRisk New York, NY
Quantitative Analyst Intern Apr 2019 – Aug 2019

- Developed web scraper to collect SOFR futures' intraday quotes from exchange websites, structured and stored in local database
- Built Qunatlib SOFR modules (ratehelper, yieldcurve, bootstrap) and created corresponding Python modules
- Collected interest rate instruments data including Fed Funds futures, Eurodollar futures, SOFR futures, FRA, Swap to bootstrap LIBOR, SOFR yield curves and model the LIBOR-SOFR spread with Multi-Curve Model using Python Quantlib
- Created long-running workflow and accomplished automations of data collection, database monitoring, data visualization and error notification system

Everbright Securities Shanghai, China
Quantitative Analyst Intern Jun 2016 - Dec 2016

- Analyzed the reversal effect on large-cap SSE 50 Index and small-cap CSI 500 Index and implemented pairs trading strategy with ETFs and designed market monitoring GUI system for triggering trading signals
- Applied Factor-Score model to the analysis of the factor style change and back-tested the half-time multi-factor alpha strategy based on industry classification enhancement and effective factor tracking

INTERESTS

- Riichi Mahjong, Poker, Billiard, Skiing, Sci-Fi

PAOLO EMILIO BARBANO, PhD
Jersey City, NJ
+1-203-936-8284
PAOLOEMILIO.PROTAGORAS@GMAIL.COM

US-Citizen. Hands-on science manager, computational scientist in technology, biomedical (Cambridge, Yale, Rockefeller universities), and financial sectors. 12+ years of experience and publications track-record in designing and leading the implementation of Signal Processing, Data Analysis and Machine Learning Algorithms for technology, biomedical, biometric, aerospace and financial applications. Specific experience advanced analytics, biometrics and wearable sensors.
Highly experienced in science-product team interactions in fast pace, dealing with ambiguity, delivering results and mentoring science teams.

Professional History

Amazon, New York NY: 12/2019-present, Senior Applied Scientist
- Leads the research, design and development of the full spectrum of Advertisement Measurement Products in close-range interaction with non-technical product team and management.
- Leads research, development and implementation in a science team, comprising Applied- and Data-Scientists, supporting the entire spectrum of Ad Measurement Products, providing full-stack Science support to Predictive, A/B testing, Statistical, NLP- and Computer Vision workflows.
- Led Technical effort and Business Implementation of Machine Learning algorithms for medical, health and financial sector AWS customers.
- Designed and Implemented Signal Processing and Machine Learning solutions for DoD Customers. Worked at customizing AWS Machine Learning Services for specific high-profile customers in the public sector.

Protagoras Technology Solutions, Arlington VA: 05/2019-11/2019, Lead Data Scientist
Led the data science effort of the firm, consulting for financial and tech industries.
- Designed innovative signal processing schemes for human posture detection from Biometric Sensors.
- Directed Engineering teams in collecting data and implements ML-based classification of Alternative Sensor Data.
- Designed Alternative Investment Strategies for Credit Trading & Bond-ETFs for a Credit Hedge Fund.

Millennium Partners, New York NY: 02/2018-04/2019, Senior Quantitative Research Scientist
Lead of Data Science Modeling and Strategy creation for two portfolios
- Designed and implemented AI-based Algorithms for Energy Commodity Trading.
- Created Traditional as well as Machine Learning-based Strategies for Trading Natural Gas and Crude Oil.

Hutchin Hill Capital, New York NY: 11/2016-02/2018, Senior Research Scientist
- Hired a team which led the effort in designing and implementing signal processing solutions for equity trading.
- Designed and Implemented a noise-robust detection/estimation algorithm for outlier Volume events.
- Created Machine Learning Algorithms Anomaly Detection in Equity Prices.

Bloomberg LP, New York NY: 09/2015-10/2016, Senior Data Scientist
- Machine Learning Workstream Leader for Bloomberg terminal LQA function.
- Designed Machine Learning framework for liquidity assessment of Corporate Bonds.
- Worked with business stakeholder and product team on its implementation in production.

Duke University, Durham NC: 09/2012-08/2015, Research Faculty
- Researcher in the biomedical and biometric field.
- The role required researching and prototyping a family of next-generation object recognition, signal processing and data mining algorithms.
- The role also required interaction with business stakeholders and discussing the implementation in the commercial space.

Yale University, New Haven CT: 08/2011-09/2012,Lecturer, Department of Mathematics
- Developed a system for the analysis of text and picture data in Medical Reports.
- Designed and developed a semisupervised algorithm for the processing of flow Cytometry data.

The University of Cambridge, Cambridge UK: 09/2009-08/2011, Senior Research Associate, DAMTP
- Actively researching (and publishing) new techniques for the rapid analysis of PET/SPECT data. Directly working with the Addenbrook hospital Imaging Center.
- Working with the MRC laboratories at Computer Vision techniques for C-elegans brain activity mapping in vivo.

Selected Publications

- S. Mao, PE Barbano, Xie, "Identification of Wireless Network Cards Based on Signatures of Wireless Traffic", ISCC2015, Larnaca Cyprus, July 2015
- N. Pelak, M. Bartlett, J. Albertson, PE. Barbano, A. Porporato, “Theoretical considerations for stochastic soil moisture dynamics and the optimal design of soil moisture sensor networks”, CUASHI 2015.
- T. Kuhn, PE. Barbano, M.L. Nagy, M. Krauthammer, “Broadening the Scope of Nanopublications”, ESWC, Semantics and Big Data Conference 2013.
- PE. Barbano, M. Nagy, M. Krauthammer, “Energy-based architecture for the classification of publication Images”, Biomedical Science and Engineering Conference, Oakridge National Laboratories, 2013.
- S. Van Vaerenbergh, I. Santamaria, PE Barbano “A Semi-Supervised Classifier Based on Connected Image Transformations”, Expert Systems with Applications Volume 40, Issue 17, 1, 7069–7079. December 2013.
- PE. Barbano, A.S. Fokas, “Multi-resolution inversion algorithm for the attenuated radon transform”, 2011 IEEE International Workshop on Machine Learning for Signal Processing.
- B. Judge, J. Brady, R. Levis, PE. Barbano, "Nonresonant Femtosecond Laser Vaporization with Electrospray Post-Ionization for In Vivo Plant Tissue Typing using a Compressive Linear Classifier", Analytical Chemistry, 2011.
- PE. Barbano, A.S. Fokas, G. Kastis, “Analytical Reconstructions for PET and SPECT employing L1 Denoising”,Proc. Conference on Digital Signal Processing, Santorini, 2009.
- S. Van Vaerenbergh, I. Santamaria, PE. Barbano, U. Ozertem, D. Erdogmus. “Path-Based Spectral Clustering for Decoding fast Time-varying MIMO Channels”, Proc. Conference on Machine Learning and Signal Processing, Grenoble, 2009.
- PE. Barbano, M. Spivak, L. Greengard, M. Flajolet, A. Nairn, P. Greengard, “A mathematical tool for exploring the dynamics of biological networks”, Proc. Nat. Ac. Sci. USA, 2007 November 21.
- Barbano PE, Spivak M, Feng J, Antoniotti M, Mishra B., “A coherent framework for multiresolution analysis of biological networks with ‘memory’: Ras pathway, cell cycle, and immune system.”, Proc Natl Acad Sci USA, 2005 May 3.
- Cao Y, Semanchik N, Lee SH, Somlo S, Barbano PE, Coifman R, Sun Z, “Chemical modifier screen identifies HDAC inhibitors as suppressors of PKD models.”, Proc Natl Acad Sci USA, 2009 Dec 22
- Ning F, Delhomme D, LeCun Y, Piano F, Bottou L, Barbano PE., “Toward automatic phenotyping of developing embryos from videos.”, IEEE Trans Image Process. 2005 Sep;14(9):1360-71.
- Stolc V, Gauhar Z, Mason C, Halasz G, van Batenburg MF, Rifkin SA, Hua S, Herreman T, Tongprasit W, “A gene expression map for the euchromatic genome of Drosophila melanogaster.”, Science. 2004 Oct 22.

Education

Laurea (BS) in Mathematics, The University of Rome, “La Sapienza”.
PhD in Mathematics, The Graduate Center of CUNY.

Programming Skills: Python, C/C++, MATLAB, AWS
Spoken Languages: Fluent: English, German, Italian, Spanish, French. Basic: Standard Mandarin

Further Technology Development Achievements

Raytheon Company, Tucson AZ: Algorithm Designer for Target Recognition/Classification and Radar/SAR Waveform design.

- Invented of a new poly-phase multi-purpose orthogonal waveform family, US6125378 (received multiple citations, including some recent ones from Google Inc., US8374218, US8385470, US8654817).
- Invented computational solutions for the automatic recognition of features for remote sensors which were inserted in commercial applicationn: knowledge was used on multiple Patent applications (e.g. US WO 2007044051 A3, WO 2006130682 A1), some of which are filed for abroad (EP 1880232 B1).

Argonne National Laboratories, Chicago IL: Computational Biology Division, developed algorithms for the detection of Anomalies in Genetic Coding data.
Hughes Space and Communications

- Created high performance multi-resolution Waveform family (US5726658A) which were inserted by Hughes Space and Communications in multiple Satellite ranging application.
- Patent No. # 5726658 was also recently cited multiple times by Boeing (see e.g. US# Pat. No. 7653156 and US Pat. No. US Pat. No. 7245599) as well as Direct TV (in US# Pat. No. 7236508, US# Pat. No. US Pat. No. 7245599) and several others.

Math Star Inc. Minneapolis, MN: Created efficient coding-modulation schemes for wireless communication.
Fast Mathematical Algorithms & Hardware, Hamden CT: Technical Consultant on electromagnetic code, Radar-waveform design, Wireless Communications

Awarded Patents

US 5726658 A - L. Auslander, PE. Barbano, U. Von der Embes, X-G Xia, R. M. Matic, “CDMA code generator employing mixing ergodic transformation”, 1998.
US 6466142 B1 - PE. Barbano, “Method and apparatus for generating families of code signals using multiscale shuffling”, 2000.

Phong Ha

Berkeley, CA | 267-506-4275 | phongtheha@berkeley.edu | linkedin.com/in/phongtheha | github.com/phongtheha

EDUCATION

University of California, Berkeley, CA — May 2023
Master of Engineering in Computer Science (EECS), Data Science concentration — GPA: 3.94
Relevant Courses: Deep Reinforcement Learning, Machine Learning, Natural Language Processing, UI Design

Temple University – College of Science & Technology, Philadelphia, PA. — May 2019
BA, Mathematics, Minors in Computer Science, Economics, Data Science — GPA: 3.90

TECHNICAL SKILLS (5 years)

Programming: Python, SQL (MySQL, Postgres), Java, C++, Matlab, STATA, HTML, Javascript, CSS
General: Data Analytics, Data Visualization, Predictive Modeling, Artificial Intelligence, Data Mining, Statistical Analysis
Tools & Frameworks: Tableau, MS Excel, PowerBI, Keras, Tensorflow, PyTorch, AWS, MLFlow, Jupyter, NLTK, Gensim, LaTeX, scikit-learn, SageMaker, PySpark, Flask, Numpy, Pandas, Git, UiPath, Agile

PROFESSIONAL EXPERIENCE

AMERICAN WATER, Camden, NJ

Data Scientist - RPA Automation Team — July 2021 – March 2022

- Assessed 40 automation opportunities with business stakeholders across 4 business units and identified 5 leading use cases for 2021 using data analytics and process mining, which would potential save millions of dollars.
- Led end-to-end implementation of Accounts Payable Reporting Automation process using UiPath, SQL, Excel VBA, automated 100% of manual effort, increased frequency from semi-daily to daily, and **saved 2000 manual hours.**

Data Scientist — June 2019 – July 2021

- Analyzed complex data sets using advanced statistical methods, predictive models, and machine learning techniques to transform raw data into actionable insights using Python, Pandas, Tensorflow, AWS Sagemaker. Leveraged these insights to provide valuable recommendation to the business and solve business problems (examples below)
- Utilized Recurrent Neural Network and other time-series forecasting models with Tensorflow and statsmodel to predict call center volumes, achieving **75% accuracy to help call center's staff planning**.
- Analyzed large datasets of 2 million customer service reviews with text mining and exploratory data analysis (EDA) using Pandas, SQL, Excel. Built deep neural networks using Tensorflow and Keras to classify reviews with 88% accuracy, saving business **80% of manual effort** required to **improve customer experience**.
- Collaborated with product team to predict user behavior on internal apps using RNN, leading to **a 98 % improvement in user engagement.**
- Deployed Human Resource Chat Bot using Dialogflow and Flask. Made accessible to all employees on Microsoft Teams by building secure REST APIs/Webhooks on AWS cloud. **Reduced HR query volume by 40%.**

BUREAU OF LABOR STATISTICS, Philadelphia, PA — August 2017 – Dec 2017
Data Analyst Intern

- Conducted research on the employment dynamics of the 5 fastest-growing occupations and industries in the past 20 years, utilizing BLS's data. Delivered comprehensive deliverables, including validated employment projection hypotheses with a 90% accuracy rate and valuable insights to inform workforce planning and policy decisions.
- Led presentation with time-series data visualizations to the Mid-Atlantic BLS Office and wrote press release.

RESEARCH EXPERIENCE

TEMPLE UNIVERSITY, Philadelphia, PA
Machine Learning Researcher — July 2018 – May 2019

- Cleaned and analyzed large text dataset consisting of 2+ millions career profiles. Performed clustering and data visualization using Tableau, matplotlib and seaborn to discover career trends in the job market.
- Developed an algorithm with Tensorflow called Iterative Mimicking (IM) that resulted in a better character-level embedding generator, capable of assigning a vector for any input strings, including OOV and rare words.
- Applied the proposed algorithm to solve the problem of **job title normalization** by representing all job titles, job descriptions in a semantic vector space. Outperformed 4 competitive baselines by 11% in accuracy.

Data Research Assistant — July 2017 – May 2018

- Leveraged BeautifulSoup and Pandas to scrape and wrangle research data from more than 200 sites, ensuring the quality and usability of data for analysis
- Assisted Economics Department Chair with data visualization and data analysis using STATA and Python

PUBLICATIONS

Ha, Phong, Shanshan Zhang, Nemanja Djuric, and Slobodan Vucetic. "Improving Word Embeddings through Iterative Refinement of Word-and Character-level Models." In *Proceedings of the 28th International Conference on Computational Linguistics*, pp. 1204-1213. 2020.

SANKET GUPTA

West Lafayette, IN | 765-476-6282 | gupta795@purdue.edu

PROFILE

- Data driven professional with 4 years of experience in marketing analytics seeking job as a data scientist.
- **Technical Skills:** R, Python, SQL, Tableau, Power BI, MS Excel, Minitab, Google Analytics, A/B Testing, ML Algorithms (XGBoost, CatBoost, LightGBM, CNN, Clustering, Regression), Genetic Algorithms.
- **Soft Skills: t**hough leadership, client communication, self-driven with good business etiquettes.
- **Core Competencies:** Marketing Analytics, Web Analytics, Machine Learning, Customer Analytics, Data Visualization.

EDUCATION

Purdue University, Krannert School of Management (GPA – 3.74/4.00) **West Lafayette, IN**
Master of Science in Business Analytics and Information Management **May 2023**

- Recipient of Merit Based Scholarship

Thapar Institute of Engineering and Technology **Punjab, India**
Bachelor of Engineering, Electrical and Electronics Engineering **May 2018**

PROFESSIONAL EXPERIENCE

ZS Associates – Customer Centric Marketing - Analytics **New Delhi, India**
Decision Analytics Associate Consultant **July 2020 – June 2022**

- Created a strong performance indicator called "Customer Prescribed Index" for Ex-US countries as part of a proof of concept leveraging **XGBoost** with an accuracy of **94%**.
- Implemented a **Next Best Action** model using **CNN and Genetic Algorithm** to create a customized customer journey for over **150K Physicians** generating an additional lift of **4% (~5MM)**.
- Identified new target physicians as part of a dynamic targeting project using **Positive Unlabeled Learning and K-Means Clustering**.
- Delivered 8+ **Marketing Mix Optimization and Attribution Modelling** projects achieving an additional revenue in the range of $**2 MM – $6 MM** by maintaining previous year's promotional budget.
- Managed **10+ client relationships** by presenting results and ensuring on-time delivery of project assignment across different countries in US, APAC, and EU regions.
- Coordinated implementation of setting up the **R-Shiny app** as an algorithm as a service for clients to run Attribution models themselves on **AWS platform**.
- Documented several R and Python codes as part of different in-house products.
- Lead teams for several projects and delivered them with results appreciated by clients.

ZS Associates – Customer Centric Marketing - Analytics **New Delhi, India**
Decision Analytics Associate **June 2018 – June 2020**

- Set up an **A/B Testing** tool from scratch to test different subject lines of a campaign and further optimize to increase engagement rate.
- Implemented various **Attribution Models** such as **Cross-Sectional, Time Series and Panel/Mixed model** spanning across consumer tactics such as banner ads and physician tactics such as sales force detailing etc.
- Built an R-Shiny Tool on R for Attribution Models with several features such as **Outlier Detection, Correlation Matrix, Ad Stocking, Transformation, Impact Assessment, Response Curves** etc.
- Standardized multiple processes on **R and Python** to make several interim processes faster.
- Collaborated with several clients in identifying solutions to different ad-hoc requests.
- Mentored junior associates on different projects and helped them understand different ML concepts.

LEADERSHIP ACTIVITIES, AFFILIATIONS, HONORS

- Student Volunteer at **Pratigya and Rotaract Club from** 2015 – 2018.
- **Certifications: AWS Cloud Practitioner**, Managing Uncertainty in Marketing Analytics, Introduction to Marketing, Data Scientist's Toolbox.