Allison Redfern

New York, NY | 412-735-3234 | amr10211@nyu.edu

EDUCATION

New York University, Center for Data Science | New York, NY *May 2024*
Master of Science, Data Science
Relevant Coursework: *Introduction to Data Science (Inference & Machine Learning), Probability and Statistics for Data Science*

University of Michigan, College of Engineering | Ann Arbor, MI *Sep 2015 – May 2019*
Bachelor of Science in Engineering, Industrial and Operations Engineering | Minor, Music
GPA: 3.8 / 4.0 Honors: *summa cum laude*, University Honors, Dean's Honor List
Leadership: Student Advisory Board, *IOE Dept Representative*
Engineering Peer Mentor Program, *Mentor*
Alpha Pi Mu Industrial Engineering Honors Society, *Tutoring Chair*
Women's Glee Club, *Social Chair*

SKILLS

Programming Languages (*Python, SQL, MATLAB, VBA*)
Statistical Software (*SPSS, Minitab*)
Data Visualization (*Tableau, Power BI*)
Database Tools *(SQL Server Management Studio, Microsoft Access)*
Data Preparation Tools (*Alteryx, Tableau Prep, Power Query*)
Optimization Software (*AMPL*)

EXPERIENCE

Johnson & Johnson

Senior Digital Product Management Specialist | Skillman, NJ *Apr 2022 - Present*

- Manage and develop Power BI application for J&J's newest and largest MedTech Distribution Center, enabling leaders to make data-driven decisions to improve inventory management, staff productivity, and customer fulfillment as the DC expands inventory
- Launched 15 new application features since starting the role in April 2022, including new data connections, visualizations, and metrics as well as application performance and reliability improvements
- Collaborate with technology teams to proactively source, cleanse, and democratize data needed for future initiatives and KPIs
- Advance organization's future data-driven initiatives by serving as Power BI Platform Lead and partnering with external consultants

Supply Chain Customer Solutions Analyst | Piscataway, NJ *Jun 2020 – Apr 2022*

- Created automated dashboards that identify opportunities to improve medical device customer supply chain performance as well as facilitate execution and measure impact of efficiency solutions
- Deployed over 200 data quality and accessibility enhancements to dashboards
- Managed and continuously improved over 20 data sources, 80 Tableau dashboards, and associated documentation
- Connected teams across J&J with data and dashboards needed to solve problems and advance projects by serving as a subject matter expert in medical device customer data
- Led Tableau trainings to increase organization's data and analytics skills shown by a 46% decrease of basic data requests in 2021
- Optimized hospital supply chain processes by collaborating with leadership from large hospital networks

Supply Chain Customer Solutions Associate Analyst | Piscataway, NJ *Jun 2019 – Jun 2020*

- Supported analytics on medical device customer metrics such as price accuracy, order consolidation, and paper invoice reduction
- Triaged and resolved data inquiries and developed tracker to collect and analyze information on team's data issues and findings
- Launched interactive data documentation and trainings to improve organization's analytical skills and data architecture visibility
- Led the Vendor Managed Inventory (VMI) program by collaborating with J&J planners and hospital buyers, leading monthly metric reviews, and redesigning the VMI SOX compliance process

Export Logistics Services Co-Op | Piscataway, NJ *Jul 2017 – Jan 2018*

- Analyzed J&J export data, consisting of over $3 billion in sales and 70,000 invoices, and reported monthly scorecard to international leadership and key business partners. Automated scorecard analytics to improve accuracy and reduce completion time by 38%
- Built data visualizations to improve shipping lane efficiency, reduce product damages, and forecast costs of process changes

Amazon

Area Maintenance Manager Intern | Phoenix, AZ *May 2018 – Jul 2018*

- Managed team of 30 maintenance technicians to ensure technology reliability at the PHX6 Amazon Fulfillment Center
- Optimized cardboard recycling process by analyzing operational data to design a conveyor system that saves $350,000 annually and decreases corrugated board disposal labor hours by 50%

University of Pittsburgh Engineering Education Research Center

Undergraduate Research Assistant | Pittsburgh, PA *May 2016 – Aug 2016*

- Investigated the impact of international experiences on the global preparedness of undergraduate engineering students by analyzing survey data and leveraging methods such as linear regression, mean comparisons, and effect sizes

Aradhita Bhandari

aradhita.bhandari@nyu.edu linkedin.com/in/aradhita-bhandari github.com/AradhitaB

Education

New York University **Sept 2022 – May 2024 (anticipated)**
Master of Science in Data Science *New York, United States*

Vellore Institute of Technology **Jun 2018 – Jun 2022**
Bachelor of Technology in Information Technology *Tamil Nadu, India*

- *CGPA: 9.09/10.0*
- Coursework: Natural Language Processing, Soft computing, Artificial Intelligence, Big Data Analytics, Data Mining Techniques, Open Source Programming, Human Computer Interaction, Information Security

Skills

- **Programming:** Python(Numpy, Pandas, Matplotlib, NLTK, Spacy, Scikit-learn, TensorFlow), SQL, R, Java, C++
- **Technologies:** Jupyter Notebooks/Labs, Spyder, R Studio, Git/GitHUb, GSuite, Microsoft Office Suite, Tableau, Hadoop

Experience

Bosch, India **Jan 2022 – Jun 2022**
Data Science Intern *Cloud and Mobility*

- Performing Aspect Based Sentiment Analysis via Topic Modelling using **NLP in Python3**
- Maintaining cloud scalability and reproducibility using **MLops techniques** and planning weekly **agile milestones**
- Creating automated process for scraping, processing, and **concealing Personal Identifiable Information** in Tweets for each month in batches
- Presenting external stakeholders at **Mahindra Group** with analytic visualizations in Tableau about the product

Blob City, India **Aug 2021 – Dec 2021**
Data Science Intern *AutoAI and AI cloud*

- Development and testing of automatic feature selection in smart "AutoAI" product and **open source package**
- Using **click-stream analysis** to identify areas in websites where users face issues or lose interest; presenting the results to external stakeholders at **Tata Starbucks** and resulting in improvement of user retention by 14%
- Analysing engine maintenance logs and critical error logs to preemptively predict failures; presenting visualizations to external stakeholders at **Delta Airlines**
- Using R, Python3, and Excel, and Tableau for Data cleaning, organization, and visualization based on stakeholder requirements

Excubator, India **Jun 2021 – Aug 2021**
Data Engineering Intern *Venture Capital & Private Equity*

- Creating algorithms in Python3 to extract relevant information from bulk data automatically
- Engineering pipelines to clean, transform and analyze data; automating Excubator's manual data handling
- Improved efficiency of data pipeline from **6 hrs to 2 minutes** per week via automated pipelines

Academic experience

Undergraduate Research Assistant **Jan 2021 – Dec 2021**
Vellore Institute of Technology *ML, Blockchain, XAI, healthcare*

- Working with Dr. Aswani Kumar Cherukuri, ex-Dean, and Dr. Balakrushna Tripathy, ex-Dean, of School of Information Technology, VIT
- Spearheading **4** different research projects in the fields of Cybersecurity, ML, Healthcare, and Blockchain technologies
- Presenting findings to professors in other universities and departments for collaboration on Healthcare related **Genetic algorithm** research
  - Aradhita Bhandari, B. K. Tripathy, Khurram Jawad, Surbhi Bhatia, Mohammad Khalid Imam Rahmani, Arwa Mashat, "Cancer Detection and Prediction Using Genetic Algorithms", Computational Intelligence and Neuroscience, vol. 2022, Article ID 1871841, 18 pages, 2022. https://doi.org/10.1155/2022/1871841

eXplainable AI in Machine Learning **Aug 2021 – Dec 2021**
Vellore Institute of Technology

- Using correlation and sampling to handle learning on highly skewed data of 2830743 rows x 79 columns with 15 classifications; using XAI to handle skewness and improving accuracy from **82% to 97%**
- Testing various set-ups for improving accuracy such as preliminary zero-day and clustering models prior to detailed classification
- Interpreting results using LIME, SHAPley values and comparing between Explainable Boosting Machines and blackbox models

Leadership / Extracurricular

Computer Society of India, VIT **Jun 2019 – July 2021**
Board Member, Ideation Head

- Spearheading project innovation and internal events
- Coordinated a team of **30 undergrad students**
- Coordinator for DevSpace (2021) - Awarded South India's Largest Student Developers' Conference (2018)

Make A Difference, NGO **Nov 2019 – Jun 2021**
Wingman

- Mentored 3 12th grade underprivileged students in soft skills, technical skills, and provided tutoring
- Worked with the students and their guardians for college applications
- Collecting and handling sensitive data for the youth throughout the year

Arda Atalik

Ph.D. Student at NYU Center for Data Science

@ arda.atalik@courant.nyu.edu 424-977-9933 416 W 56TH ST, NEW YORK, NY 10019 ardaatalik.com

Education

Doctor of Philosophy in Data Science

NYU Center for Data Science

Sept. 2022 – Ongoing New York, NY

- Advisors: Prof. Sumit Chopra, Prof. Daniel Sodickson, Prof. Kyunghyun Cho
- Center for Data Science Fellowship

Master of Science in Electrical and Electronics Engineering

Bilkent University

Sept. 2019 – Aug. 2022 Ankara, Turkey

- GPA: 4.0/4.0
- Advisor: Prof. Orhan Arikan
- Thesis: Innovation Based Transmission in AI-Enabled Sensor Networks for 6G IoT Scenario
- 5G Fellowship and The Scientific and Technological Research Council of Turkey Fellowship

Exchange Study in Electrical and Electronics Engineering

École Polytechnique Fédérale de Lausanne (EPFL)

Jan. 2017 – Aug. 2017 Lausanne, Switzerland

Bachelor of Science in Electrical and Electronics Engineering

Bilkent University

Jan. 2014 – Aug. 2018 Ankara, Turkey

- GPA: 3.97/4.0
- Comprehensive Scholarship and High Honor standing in all semesters
- Academic Excellence Award based on top 1% ranking

High School Degree in Natural Sciences and Mathematics

Meram Science High School

Sept. 2009 – June 2013 Konya, Turkey

- GPA: 93.19/100, ranked 320^{th} in the national University Entrance Exam amongst 1.8 million high school graduates

Work Experience

5G Research Engineer

Vodafone

Sept. 2019 – Nov. 2020 Ankara, Turkey

- Vodafone-Bilkent joint project on computationally efficient beam-forming techniques for 5G and beyond networks

Graduate Research Student – EDIC Fellow

École Polytechnique Fédérale de Lausanne (EPFL)

Sept. 2018 – Aug. 2019 Lausanne, Switzerland

- Worked on the remote source coding problem and compute-forward multiple access for Polar codes in the Laboratory for Information in Networked Systems, headed by Prof. Michael Gastpar

Internships

- **2016 – Summer Intern** at Bilkent Univ. NANOTAM Xilinx Virtex 7 VC707 FPGA Programming
- **2018 – Summer Intern** at Polaran
  Studies on Polar Codes and LDPC Codes for the 5G NR under the supervision of Prof. Erdal Arikan

Selected Publications

Differential Entropy of the Conditional Expectation under Additive Gaussian Noise (first author)

IEEE Transactions on Signal Processing

The Price of Distributed: Rate Loss in the CEO Problem (first author)

2022 Conference on Information Sciences and Systems (CISS)

Towards goal-oriented semantic signal processing

Elsevier – Digital Signal Processing

June 15, 2021

Variations on Hammersley's interacting particle process (first author)

Discrete Mathematics Letters

June 19, 2021

Radar Antenna Selection for Direction-of-Arrival Estimations (first author)

2021 IEEE Radar Conference (RadarConf21)

June 18, 2021

Atlanta, USA

Computer Skills

Basic Knowledge: VERILOG, Mathematica, Maple
Intermediate Knowledge: C++, JAVA, Julia, VHDL
Advanced Knowledge: C, Python, MATLAB

Areas of Interest

- Mathematical Foundations of Machine Learning
- Machine Learning for Healthcare
- Communication Systems
- Signal Processing
- Information Theory

More Information

Google Scholar
`https://scholar.google.com/citations?user=u6uEIDkAAAAJ`

Exams

TOEFL IBT (09/04/21)

- Reading: 28/30
- Listening: 28/30
- Speaking: 23/30
- Writing: 26/30

GRE (11/08/17)

- Verbal: 146/170
- Quantitative: 170/170
- Analytical Writing: 3.5/6

Meiyu (Emily) Li

Cell: 201-234-2606 | **Email**: ml8457@nyu.edu | https://www.linkedin.com/in/meiyu-li-58a10b156/
Add.: 235 Grand St., Jersey City, New Jersey. Zip code: 07302

Working as a **Data Analyst** in 3 jobs and 5 research projects, I am highly experienced in cooperation as well as leadership, having helped groups **develop statistical analysis** and gain more insights. I am a **team-player, risk taker, and positive-minded individual** who looks forward to bringing change in my community. I am seeking a data analyst role in technology field, which will enable me to contribute my skills and enrich my knowledge.

EDUCATION

New York University, Center for Data Science, NY **09/2022-05/2024**
Master of Science in Data Science

- Courses to be taken: Intro to Data Science, Probability and Statistics for Data Science, Linear Algebra and Optimization, Machine Learning (expected), Big Data (expected), Database Management (expected), Natural Language Processing (expected)

University of California Santa Barbara, Department of Applied Statistics | Department of Mathematics, CA **09/2018-09/2021**
Bachelor of Science in Statistics and Data Science | **Bachelor of Science in Mathematics** **GPA: 3.92/4.0**
Honors: Dean's Honors for Each Quarter in College of Letters and Science

- Courses: Machine Learning, Big Data with PySpark, Data Visualization, SAS, C++, Regression Analysis, Survival Analysis, Logic, Stochastic Process,Probability, Calculus, Linear Algebra, Real Analysis, Topology, Abstract Algebra, Design of Experiments

SKILLS & COMPETITION

- **Skills**:
  - **Proficient: R, Python,** SAS, Excel, Word
  - Intermediate: Java, C++, Octave
  - General: Html, JavaScript, SQL
- **Competition**: Gold Medal in UK Mathematics Competition of 2017, Global Top 5% in Waterloo Math Contest of 2018

PROFESSIONAL EXPERIENCE

YiduCloud (Beijing) Technology Co. Ltd. **01/2022-07/2022**

- Worked as a **full-time medical data analyst** with **R** and **Excel** for clients in pharmaceutical companies, hospitals, medical labs
- **Made survival analysis** *(eg. KM Model, Cox PH Model)***, descriptive statistical analysis, data cleaning, preprocessing and wrote statistical analysis plans in 5+ projects**

Tencent (Shenzhen) Computer System Co. Ltd. **09/2020-11/2020**

- Worked as an intern of Platform and Content Group for the operation of Xiaoe Pinpin, an online shopping platform
- Recorded the data such as the daily/monthly sales, recommended products' information, and **made market analysis reports with Excel and Python** every week

Statistics and Applied Probability Department of UC Santa Barbara **09/2020-09/2021**

- Worked as a **teaching assistant** of the undergraduate course, **Understanding Data**
- Assisted the professor in grading homework, **taught elementary statistics involving Python programming and basic statistics theories**

PROJECT EXPERIENCE

Independent Research on Predicting Beijing's House Price with R **01/2021-06/2021**
Advisor: Professor Tomoyuki Ichiba

- Analyzed and predicted the housing price in Beijing **by Linear, Ridge, Lasso Regressions, Classification, Time Series** and gleaned more data information by **Web Crawler**

Research on Users of Spotify with PySpark **01/2021-03/2021**
Advisor: Professor Adam Tashman

- **Predicted** the **popularity** of Spotify **songs** and **classified** song titles in a group with 2 classmates and made report and presentation
- Made **analysis with feature transforms** by **comparing with Regressions, Decision Tree, Random Forest, GradientBoosting, Naïve Bayes**

2016 US Presidential Election Analysis with R **09/2020-12/2020**
Advisor: Professor Zhijian Li

- Explored the demographics' role and voter behaviors' impact on presidential election with 3 classmates
- Made analysis through **Principal Component Analysis, Boosting, Random Forest, Logistic Regression and Support Vector Machine**, etc.

Data Analysis on Portugal Hotel Booking Demand and Revenue with Python. **03/2020-06/2020**
Advisor: Professor Alexander Franks

- Described the Portugal hotel's revenue factors, including price, cancellation rate, booking time, customer behaviors with 1 classmate
- Made **exploratory data analysis, Principal Component Analysis, and developed multilinear regression model**

SHIH-LUN (ALLEN) HUANG

NEW YORK, NY, 11373 | (929) 544-7817 | allen.sl.huang@gmail.com | LinkedIn

EDUCATION

New York University, Center for Data Science — June 2024
Masters in Data Science

National Tsing Hua University, Taiwan — June 2021
Bachelor of Quantitative Finance
Elective courses in data science

SKILLS

Programming: **Python (Intermediate)** (GitHub), **SQL (Expert)**, MATLAB

Technical Application: *PCA, Regression,* ***Optimization****, Gradient Descent,* ***Machine Learning, Text-Mining,*** *Web Crawling, Data Visualization, Corporate Finance, Derivative Pricing*

EXPERIENCE

Pinkoi, ***Data Scientist Intern*** *(Full-time)* — July 2021 – Feb 2022

- **Analyzed and modeled (Regression** *and* ***Machine Learning*** methods**) quantitative** and **qualitative data** in **AWS database** for user **behavior research**
- **Conducted experiment A/B testing**, and **product optimization** for e-commerce web\app platforms
- Constructed **user road map** to improve understanding of different phases of user behavior and increase **user conversion rate**
- **Supported product feature designing** and key objective goal planning with a 12-person team consisted of engineers, product designers, and a project manager
- **Administered** and **taught** a series of **SQL** and **data visualization sessions** to **60 plus colleagues** from sales team and product team

Department of Quantitative Finance at NTHU, ***Research Assistant*** — June 2020 – July 2021

- **Processed stand-alone CSR reports and MSCI KLD scores** in **Python** by conducting **web scraping**, **data cleansing**, and **textual analysis**
- **Reviewed** and **gave presentations** on **over 20 academic papers** regarding topics such as, **NLP analysis** in **finance**, and **influence of CSR performance** on firm financial performance

PROJECTS

Hedging Climate Change News (Replication) — Feb. 2021 – June 2021

- **Constructed a proxy index** to capture climate change innovations by **conducting text anaylsis and data mining** on over 4 years of daily news from Wall Street Journal, 15 whitepapers and 55 glossaries on climate change
- **Built a mimic portfolio by modeling volatility** of over 4000 securities in relation to climate change index to hedge climate change innovations

Does firm performance improve by following the reporting guidelines of SASB materiality map? — March 2021 – July 2021

- Researched on the methods of **CSR indexes** including SASB, MSCI KLD, sustainable index, and more
- **Analyzed** changes in the proportion of different CSR categories **with text mining** on stand-alone CSR reports

LEADERSHIP AND ACCOMPLISHMENTS

NTHU Kenya International Volunteer Group, ***Public Relations*** — Oct. 2018 – Sep. 2019

- **Raised $70,000 and 250 computers**, and helped promote sponsors, also have experience receiving company executives
- **Acquired official contract** with **education authorities in Kenya** to support our volunteering work

2019 Kenya ICT Education Workshop, ***Director/Host*** — Oct. 2018 – July 2019

- Directed an ICT education workshop for about **60 participants** including teachers, university professors and CS major students to discuss ICT education issues in Kenya at Dedan Kimathi University in Nyeri, Kenya

Wang (Joe) Jiang

525 West 52nd Street New York, NY 10019
929-319-1562| wj596@nyu.edu

EDUCATION

New York University, Center for Data Science **New York, NY**
MS in Data Science Sep 2022 - May 2024
Relevant Coursework: Probability and Statistics for Data Science, Programming for DS, Intro to Data Science

Fordham University, Gabelli School of Business **New York, NY**
BS in Global Finance and Business Economics, GPA: 3.85/4.0 Sep 2018 - May 2022
Relevant Coursework: Programming with Python, Database Systems, Structures of Computer Science, Information Systems, Statistical Decision-Making, Linear Algebra

PROFESSIONAL EXPERIENCE

American International Group **New York, NY**
Data Analyst, Apprenticeship March - Nov 2021
- Conducted deep research and summarized industry reports of housing demand, airline industry, and credit card spending after the pandemic
- Applied trending analysis on time series net migration data of major metropolitan and micropolitan areas from Census Bureau to compare the influence of pandemic
- Visualized the analysis using Excel charts and graphs and presented directly to Chief Economist

Suzuki Capital LLC **New York, NY**
Business Analyst Intern July - Aug 2016
- Conducted research based on redevelopment proposal on investment overview, financial projections, neighborhood, comparable, teams, and market study on 106 Franklin Street, New York and 4105 29th Street, Long Island City
- Summarized and presented the findings to C-level management

PROJECT EXPERIENCE

Trading Portfolio Data Analysis, bootcamp July 2022
- Evaluated the normalized price and analyzed the performance of 5 U.S stocks based on their return, volatility and sharpe ratio in python
- Optimized the portfolio of stocks based on the statistical analysis of their covariance and correlation matrix
- Generated the efficient frontier for portfolio of stocks by applying monte carlo simulation

Housing Price Projection, bootcamp June 2022
- Performed data cleaning and visualization for San Francisco historical housing data in python
- Achieved feature engineering to analyze and select the important features for modeling
- Established different regression model for predicting house prices and evaluated the performance of the model based on different metrics

Fordham Ram Van Scheduling Database System, Fordham University Sep - Dec 2021
- Designed an Entity/Relationship model with a set of normalized table and report for the proposed Ram Van Scheduling System
- Constructed a relational database with interactive data structures in PostgresSQL and transformed data models into database with specification
- Created and implemented relevant SQL DML statements and REST API in support of the system

Consulting Cup Challenge Competition, Fordham University Sep - Dec 2019
- Conducted a survey to collect feedback about major airlines in the U.S
- Performed sentiment analysis of customer satisfaction index using Pandas DataFrame and NLTK Vader
- Built a decision tree model with 0.91 accuracy score to predict the most critical issue of United Airlines' customer experience
- Evaluated United Airline's stock performance against its competitors and visualized the analysis using Matplotlib
- Collaborated with team to create a presentation and pitched idea to panel of experts

Kathleena **Inchoco**

(+1) 203-572-3391 | ki2130@nyu.edu | linkedin.com/in/k-p-inchoco/

Education

New York University Center for Data Science *Aug 2022 - Present*

M.S. Data Science, Concentration: Mathematics and Data *New York, NY*

- Coursework: Introduction to Data Science (Inference and Machine Learning), Probability and Statistics, Programming for Data Science

Wellesley College *Sep 2014 - June 2018*

B.A. Economics and Minor in Mathematics *Wellesley, MA*

- Coursework: Econometrics, Microeconomics, Macroeconomics, Multi-variable Calculus, Linear Algebra, Abstract Algebra, Physics

Massachusetts Institute of Technology *Sep 2017 - May 2018*

wellesley-MIT Exchange Program *Cambridge, MA*

- Coursework: Options and Futures Markets (15.437), EnActing Leadership (15.282), Undergraduate Research Opportunities Program (URG.14)

The London School of Economics and Political Science *Sep 2016 - June 2017*

General Course Student *London, UK*

- Coursework: Principles of Finance, Economics in Public Policy, International Organizations, Philosophy of Economics

Experience

Morgan Stanley *Mar 2022 - Aug 2022*

Macro Research: G10 FX Strategist, Associate *New York, NY*

- Recommended trades to FX research team, sales/trading, and debate ideas related to foreign exchange markets research.
- Worked on client requests related to macroeconomic themes in foreign exchange and forecasting currency valuations.
- Wrote market summary on fixed income market developments around foreign exchange currency movements with a focus on Canada, New Zealand, and Australia.

The Federal Reserve Bank of New York *Sep 2020 - Mar 2022*

Monetary Policy & Markets: Senior Markets Analyst *New York, NY*

- Analyzed Policy and Expectations Survey data and Bloomberg markets data in R and SQL for SOMA Manager Lorie Logan's desk briefing report.
- Led charts and tables analysis in R and Excel for financial markets desk briefing report to the Federal Reserve Board each FOMC cycle.
- Liaised across different staffs in the Markets Group, manage distribution of data resources, and data visualizations analysis for use in senior level briefings for the Federal Reserve System, foreign central banks, think tanks, and other public research institutions.

The Brattle Group *Jul 2018 - Aug 2020*

Economic Consulting: Senior Research Analyst *New York, NY*

- Analyzed historical FX data in R and produced remediation results to banking client related to 800+ transactions in litigation case.
- Conducted valuation analysis in Excel and extensive document review of corporate financials for health insurance M&A litigation case.
- Analyzed healthcare claims data in SQL and R for expert report studying the incidence of NICU admission under various physicians against industry NICU rate admission for healthcare claims litigation case.

Extracurricular

Women in Data Science Club at NYU Center for Data Science *Aug 2022 - Present*

Events Committee *New York, NY*

- Synthesize event ideas and execute event planning for the Women in Data Science initiative at NYU Center for Data Science.

Rewriting the Code *Sep 2022 - Present*

Member *New York, NY*

- Network with other women in Tech to find community in a space that encourages women to pursue SWE and data science roles.

Wellesley College Alumnae Association *Jun 2018 - Present*

Class President *New York, NY*

- Maintain finances, fundraise, and collaborate with 2018 class council to support sustainable platform of collegial connection to the College.

Honors & Awards

2022 **DeepMind Scholarship**, $140,00 full scholarship, tuition and fees, equipment stipend, academic conference travel, and mentorship award sponsored by DeepMind to 2 MS/PhD Data Science students at NYU.

Skills & Licenses

Technical Skills Bloomberg Terminal, R, Excel, SQL, Python. Working knowledge of C.

FINRA Securities Broker Certifications Security Industry Essentials (SIE) Exam, Series 7, Series 63

Keya Shukla

Jersey City, NJ, 07306 | +1 575 295 3496 | ks6875@nyu.edu

EDUCATION

New York University, GSAS, Center for Data Science — **New York, NY**
Master of Science, Data Science — *May 2024*

National Institute of Technology, Delhi — **New Delhi, India**
Bachelor of Technology, Computer Science and Engineering — *May 2021*

TECHNICAL SKILLS

Coding Languages: Python (libraries: numpy, pandas, matplotlib, Plotly, seaborn, scikit-learn), SQL, Java, C/C++, JavaScript, HTML, CSS
Frameworks: Tableau, Tensorflow, Keras, PyTorch, Flask, Django, Git, Google Colab, Jupyter Notebooks, Microsoft Office, Google Suite, Slack

WORK EXPERIENCE

Volunteer ML Engineer **at Omdena, Pennsylvania Chapter** — **June 2021 - July 2021**

- Collaborated with 10+ engineers on an open-source project to build a predictive model that determines the success rate of startups.
- Performed exploratory data analysis using pandas, seaborn and Plotly.
- Collaborated with model building team to create models for HHI score (AdaBoost Regressor), employee growth percentage (Random Forest Regressor), final status of startup (Random Forest Classifier).

Data Analytics Intern **at JPMorgan Chase & Co.** — **August 2020 - September 2020**

- Implemented Perspective open-source code to visualize two historical stock feed correlations in a web app for users to effectively judge stock performance and identify under and over-valued stocks thereby monitoring for potential trade opportunities.
- Created a UX solution to visualize live and historic data cleanly in a single chart instead of multiple workstations, making visualization and analysis more efficient.
- Performed web-scraping in Python to help the firm download public data, including over 10,000 company descriptions and stock quotes, enriching internal data and increasing research efficiency by over 50%.

Data Science Intern **at Exposys Data Labs, Bangalore** — **July 2020 - August 2020**

- Analyzed data quality of a medium sized company for a three-month period to determine target customers that drive most value for the company.
- Performed RFM analysis in Python to analyze customer value in database and direct marketing to identify the firm's top clients based on nature of spending habits and generated insights on customer churn and renewal rates from data tables with 100M rows in SQL.
- Prepared reports interpreting customer behavior, market conditions, market results, trends, and investment levels.

Web Development Intern **at Altro Smart Inc., Pune** — **November 2019 - January 2020**

- Developed front-end of web portal interface for company's smart lock device to control lock settings, manage guest and member lists and grant access permissions.
- Collaborated with Senior Web Developer to develop the back-end of admin user interface.

ML and IoT Intern **at Bolt IoT, Bangalore** — **May 2019 - July 2019**

- Created a light intensity monitoring device using LDR sensor that collects and sends data to Bolt Cloud to determine optimum place for plants in greenhouse through a dynamic graph by using polynomial regression.
- Developed an intruder alert system for chemical storage units that detects sudden change in temperature using anomaly detection algorithm.

INDEPENDENT PROJECTS

- **Disease Detection and Diagnostic Report Generation Using Deep Learning**
  Developed a VGG-19 classifier in Python to detect the presence of among 14 diseases in a chest x-ray and integrated it with a report generation model consisting of dual LSTM and attention mechanism to perform image captioning.
- **Market Basket Analysis**
  Discovered patterns through data analysis between products and spending habits of customers in a store dataset by finding out the lift, support, and confidence scores through association rule mining Apriori Algorithm.
- **Human Activity Recognition from Smart Phone Data**
  Classified sequences of accelerometer data recorded by existing sensors in smart phones into well-defined movements, decomposed into simpler gestures and performed activity recognition using two-layer Hidden Markov model.

LEADERSHIP EXPERIENCE

President **at 180 Degrees Consulting, NIT Delhi Branch** — **May 2019 - May 2021**

- Transformed branch marketing and branding strategy through social media which resulted in a 93% boost in engagement and response scores based on reach data.
- Signed 3 startup companies and 1 NGO on projects related to marketing, fundraising, business strategy, data analysis, increasing acquisition statistics by 74% compared to previous year.
- Ranked among top performing branches in APAC region.

LI JU

lj1332@nyu.edu | +1(217)819-0306 | www.linkedin.com/in/li-ju-8b73a5192/

EDUCATION

New York University, Center of Data Science — **New York, NY**

Master of Data Science — Sep 2022 – May 2024(Expected)

University of Illinois at Urbana-Champaign, College of Liberal Arts and Science — **Champaign, IL**

*Bachelor of Science in Statistics and Economics with Minors in Computer Science and Informatic*s — Aug 2019 – May 2022

- GPA and Honors: 3.96/4.00; Magna Cum Laude; Distinction in Economics and Statistics Departments

Nankai University, College of Economics — **Tianjin, China**

Bachelor of Arts in Economics — Sep 2017 – June 2019

- GPA:87/100; Teaching Experience: Course Assistant for Calculus I

PROFESSIONAL EXPERIENCE

Tencent Technologies Co.Ltd — **Shenzhen, China**

Data Scientist Intern in WeTV Team, Platform & Content Group — Sep 2021 –Dec 2021

- Developed **Logistic Regression**, **Random Forest**, and **Gradient Boosting Tree** models with **Pyspark** to predict customer churn, achieving AUC of 80%
- Automated the machine learning pipeline to perform daily training and prediction to help inform user retention strategies
- Implemented an ETL(Extract, Transform, Load) pipeline to load video data using **Pyspark**, **pandas** and **Airflow**; improved data update frequency from daily to hourly by optimizing table partitioning
- Created an interactive dashboard by internal **BI system** to display **SQL-calculated** metrics quantifying users' activeness and retention

ShareIt Technologies Co.Ltd — **Beijing, China**

Data Analyst Intern in ShareIt App Data Analysis Team — May 2021 – Aug 2021

- Developed key metrics quantifying the performance of file transferring and tool functions in ShareIt app, and created **Tableau** dashboards to extract insights based on multi-dimensional data pulled by **SQL** query
- Analyzed the cause of abnormal fluctuations in user activities through hypothesis testing; recommended data-driven solutions to product managers to improve the performances of mobile application
- Researched users' file transfer behaviors on Xiaomi cellphone in the Indonesia marketplace to identify the causes of file transferring failure, and presented improvement opportunities to the engineering team

Intel Corporation — **Shenzhen, China**

Supply Chain Business Analyst Intern — Dec 2020 - Feb 2021

- Coordinated with sales agents and manufacturers to maintain supply chain logistics, meet customer demand, and achieve revenue goal
- Leveraged demand prediction and stocks information in SAP system to distribute products in the Chinese market

SELECTED PROJECTS

Image Classification – A CNN-based Deep Learning Project for Classifying COVID-19 Diagnosis Result — **May,2021**

- Developed a **CNN** (convolutional neural network) model in **Keras** to predict COVID-19 diagnosis using over 2000 chest x-ray images from Kaggle, achieved 71% accuracy

Heart Disease Diagnosis – A Machine Learning Project for Classifying the type of Heart Disease — **Dec,2020**

- Conducted exploratory data analysis, data cleaning, data transformation on human health indices using **pandas** and **seaborn**
- Trained multiple ML models (**KNN, decision tree, random forest, boosted models)** with hyperparameter grid search to predict the level of severity of heart disease; achieving accuracy at 81%

A Database Website Application for Song Search, Rating & Recommendation — **July 2020 – Aug 2020**

- Led a group of 4 to build a database website with **CRUD** functions including querying songs, rating songs, adding songs to song lists, recommending friends and songs to users; the website uses **HTML, CSS** for frontend, **PHP,SQL** for backend, **matrix-factorization** technique for recommender system

Technical Skills

- Languages: Python, SQL, R, Java, HTML, CSS, PHP, MongoDB, Cypher Query Language
- ML Toolkits: Scikit-Learn, PyTorch, TensorFlow, SciPy, NumPy
- Data Visualization: Matplotlib, Seaborn, Tableau, R Shiny, Plotly, Dash, ggplot
- Software: Microsoft, Airflow, SAP

Manli Zhao

manliz@umich.edu | (734) 450-6217 | Jersey City, NJ | Linkedin

EDUCATION

Master** of **Science** in **Data Science, New York University *Dec. 2023*

Bachelor** of **Science** in **Data Science, University of Michigan *Aug. 2019–May 2022*

GPA: **3.84/4.0**; **University Honors** 2019, 2020, 2021; James B. Angell Scholar

- Relevant Coursework: ***Data Mining, Machine Learning, Database Management Systems, Web Systems, Probability Theory, Linear Algebra, Applied Regression Analysis***

RELEVANT SKILLS

- **Programming Languages:** Python | SQL | R | C++ | Java | HTML | CSS | JavaScript
- **Libraries:** NumPy | Pandas | Scikit-Learn | Matplotlib | Seaborn | Pytorch | SciPy
- **Software:** Jupyter Notebook | Google Suite | Git | SPSS | LaTeX | Stata | Tableau | Microsoft Access

PROFESSIONAL EXPERIENCE

Meituan.com Beijing, China

***Business Analyst** Intern in **Digital Advertising** Department* *May 2021–Aug. 2021*

- Applied and tuned **optimized K-means model** with **Random Partition initialization** to **categorize** city markets based on business potential features, forming clear breakdown to guide business development.
- Streamlined and automated **Ads performance analysis report** to dynamically monitor monetization efficiency, saving team **8 hours** manual work per week.

Shopee.com Shenzhen, China

***Analyst Intern** in **Strategy** and **Market** Department* *Jul. 2020– Sep. 2020*

- Analyzed questionnaires and **initialized hypotheses** on **correlation between cover image and sales performance**.
- Conducted **A/B testing** to evaluate page views and **click-through-rate** of cover images in differing styles, resulting in **20% boost in CTR** and **25% increase in sales**.
- Performed **hypothesis testing** and calculated p-value, compared it with predefined alpha to show the statistical significance between test groups to validate the A/B test results.

PROJECT EXPERIENCE

Text Sentiment Classifier to monitor Reddit comments *Jan. 2022–Apr. 2022*

- Transformed documents into **feature vector**; trained **kernelized SVM** model and fine-tuned the model using **grid search**; selected model using **5-fold cross validation** under Accuracy, AUROC performance.
- Improved binary **bag-of-words** model using **TF-IDF** score; extended features using **Next Word Negation** technique.

Image Classification: Classify dog images by breed *Jan. 2022–Apr. 2022*

- Implemented and trained **deep neural network** of own designed; using **early stopping** to prevent **overfitting**.
- Visualized what **CNN** has learned using **Grad-CAM**, identified noisy background feature was being learned by model.
- Reused the **pre-trained model** trained from larger dog dataset to initialize weights; **Augmented data** by applying **rotation** and **grayscale**, and final model achieved better accuracy and learned dog related features.

Instagram Clone: Full Stack Web App *Jan. 2022–Apr. 2022*

- Developed a dynamic Instagram clone that utilized **Python Flask** backend**, SQLite database. ReactJS** frontend and **AJAX** calls to custom **JSON-based REST API**.
- Created **shell scripts in bash** to allow programs run easily; deployed production build to **AWS EC2 instance**.

Creating Social Media ETL pipeline *Jan. 2021–Apr. 2021*

- Designed **relational database** to store metadata for a fictional social media platform using **Oracle DBMS**.
- Utilized **Java** program to connect database using **JDBC**; executed queries and placed results in **Java data structures.**
- Extracted records in **Oracle database** to JSON and load to **MongoDB**; calculated aggregate value using **MapReduce**.

Machine Learning Analysis in Financial Applications *Jan. 2021–Apr. 2021*

- Constructed models to predict 10-minute **forward return** given historical minutely prices of three assets over one year using **R**, utilizing **linear regression, KNN, Ridge, Lasso**.
- Improved model by deriving new features then running **Principal Component Analysis**, increasing **out-of-sample correlation** between prediction and true response by **6%.**

Ming (Jerry) Gao

Tel: (201)-377-8408 | Email: minggao077@hotmail.com | Linkedln: http://www.linkedin.com/in/ming-jerry-gao

EDUCATION

New York University, ***Center for Data Science*** New York City, NY
Master of Science in Data Science Expected Graduation 05/2024
Relevant Coursework: Machine Learning, Deep Learning, Natural Language Processing with Representation Learning

Yunnan University, ***School of Mathematics and Statistics*** Kunming, China
Bachelor of Engineering in Data Science and Big Data Technology | GPA: 3.83/4 09/2018-06/2022
Relevant Coursework: SQL, Data Mining, Data Visualization, Big Data Exploratory Analysis, Big Data Preprocessing

SKILLS

- **Programming**: Python (Scikit-Learn, Pandas, Numpy, Matplotlib, Seaborn, etc.), R (dplyr, caret, ggplot2), SQL
- **Data Science:** Data Analysis, Machine Learning, NLP, Data Visualization, Statistics, Web Scraping, A/B Testing
- **Other tools**: PowerPoint, Excel, Jupyter Notebook, Pycharm, Rstudio

PROFESSIONAL EXPERIENCES

Henan Junyou Digital Technology Co., Ltd. Zhengzhou, China
Data Scientist Intern 02/2022-04/2022

- Compared the similarity between Jingdong and Taobao category names and matched the Jingdong category to the Taobao category with the highest similarity layer-by-layer using the **Alibaba Cloud word vector API interface** and **cosine similarity**
- Applied the **SMOTE algorithm** to process the imbalanced **4M+** data of Jingdong in a certain month
- Constructed the **logistic regression**, **Naive Bayes**, **AdaBoost**, and **XGBoost classifiers** to predict price reduction for products, and found the optimal model with high interpretability and high accuracy of 87.6%
- Provided the company with a **project example** as it was the company's **first project** to use machine learning for data analysis during the transformation process

RELEVENT PROJECTS

Bank Credit Card Customer Churn Warning Based on Multi-class Logistic Regression Model 02/2022-05/2022

- Applied the **K-prototype clustering** algorithm to cluster **10k+** credit card customers data, dividing customers of a bank into three categories: loyal, churn and potential churn
- Conducted **One-Hot encoding**, performed **feature engineering** based on the Extra-Trees model
- Applied One-vs-Rest method to construct a **three-classification logistic regression** model, achieving 99.67% accuracy
- Explored the factors affecting the churn status of different credit card customer groups and delivered **recommendations** on customer churning prevention and improvement of customers' loyalty

Research on the Accurate Discriminative Model for Heart Disease Patients 06/2021-07/2021

- Preprocessed **3k+** data, applied a C4.5 **decision tree** model, a **back propagation neural network** model, and a **logistic regression** model to construct different discriminant models for heart disease patients using **Python**
- Compared the 3 models and reached 73.33% accuracy via the logistic regression model, explored the factors affecting the likelihood of heart disease, which could improve the efficiency of predicting heart disease (**by 30%**)

Natural Language Processing (NLP) Text Classification 12/2020-01/2021

- Led a team of 3 to scraped **1k+** pieces Covid-19 news in Chinese using **Python**, and labeled the news texts manually
- Segmented the texts using **jieba** library, removed non-text characters and stop words
- Performed a **TF-IDF** model and a **PCA** model for text feature extraction
- Constructed **AdaBoost** models and **XGBoost** models and optimized models with fine-tuning hyperparameters
- Compared the models, found that a certain AdaBoost model was the optimal model for classifying the scraped news with accuracy of 80%

Predictive Analysis of Car Dealer A's Customer Churn Warning and Car Return to Factory 05/2020-06/2020

- Preprocessed **430k+** data, used **Python** to connect the **Microsoft SQL Server** to store the data, and constructed **discriminant analysis**, **logistic regression** and **k-nearest neighbor** models to predict customer churn
- Constructed **OLS Regression**, **Ridge Regression**, **Lasso Regression** and **Principal Component Regression** models to predict the car dealer A's car return situations to the factory
- Compared the models and applied the optimal ones, founding the churn rate of the car dealer A's customers as 32.81%
- Explored the specific factors affecting the customer churn and delivered **recommendations** on reducing customer churn rate

Nikitha Kasipati

New York, NY | mk8463@nyu.edu | LinkedIn | (917)-657-0958

Energetic and passionate Data Science graduate student eager to expand theoretical and practical foundation in areas such as Machine Learning and Natural Language Processing. High-potential candidate, poised to effectively deliver on key goals and develop a 360-degree view of challenges. Valued team player, fostering diversity, exhibiting effective leadership and communication skills. and adhering to the highest ethical standards.

Education

Master's in Data Science
(Relevant courses: Probability and Statistics for Data Science, Optimization and Computational Linear Algebra)
***New York University (NYU)**, New York* | *May 2024*

B-Tech in Computer Science & Engineering
Vellore Institute of Technology (VIT)**, Vellore, India* | *June 2022* | ***CGPA: 9.20

Problem Solving and Technical Support Skills

Programming Languages – Python ● SQL ● C, C++ ● Java ● R Programming ● Angular.js ● HTML ● JavaScript ● PHP

Libraries & Frameworks – Numpy ● Pandas ● Matplotlib ● Seaborn ● Scikit-Learn ● Keras ● TensorFlow ● Node.js

Professional Experience

IT Intern **Jan 2021- Mar 2021**
World Wide Commercial Ventures Limited (WWCVL), Lagos Nigeria

- Created Power BI dashboard of sales reports of 10 years, to give better insights by geography, product grouping, category, by principal and analyze trends accordingly.
- Engineered various visualization concepts using DAX functions: deployed performance indicators like Value Growth by Product & Performance Indicator which depicted progress of sales trend of current year as compared to previous years.
- Designed a Unique Products Tracker deploying Python, Pandas & PostgreSQL, with login-page and password through Angular JS. Developed jQuery script to make search bar to filter elements & track required product.

Business Analyst Intern **Sep 2021- Dec 2021**
Scale Global, Hyderabad, India

- Tasked with scraping by parsing content from e-commerce websites using the Beautiful Soup library to get detailed product information in the skincare category.
- Formatted unstructured and semi-structured sales data from disparate sources using Python to analyze trends in sales, seasonality and geographic spreads.
- Designed personalized market strategies and optimized Amazon advertising to better target customers by doing extensive product research and locating high-volume keywords on Amazon using Helium10.

Publications

- Rajkumar, Nikitha, Jantwal "Cloud hosted ensemble Learning based rental apartment price prediction model using stacking technique" *(In Press) Deep Learning Research Applications for Natural Language Processing* (Chapter 15) IGI Global 2022

Leadership Experience

Toastmasters International Club **Dec 2018-May 2021**
Roster member, Vice President of Membership (2019-2020)

- Lead anchor for GRAVITAS 2019 - University's technical fest, one of the biggest in India.

Skillship Foundation **June 2020-June 2021**
Technical Lead

- Organized workshops to mentor and provided assistance to women and children in need of technical skills.

NUO LEI

(551) 344-2542 | nuo.lei@nyu.edu | https://www.linkedin.com/in/nuo-lei/

EDUCATION

New York University, Center for Data Science**,** New York, United States *Expected May 2024*

- *Master of Science in Data Science* *GPA: Pending*
- **Expected Coursework**: Machine Learning, Deep Learning, Natural Language Processing, Database Systems

University of International Business and Economics, Beijing, China *September 2018 - July 2022*

- *B.A. in Financial Mathematics* *GPA: 3.8/4.0*
- **Relevant Coursework**: Linear Algebra, Probability and Statistics, Time Series Analysis, Python Big Data Analysis, Data Mining and Statistical Learning, Optimization, Mathematical Modeling, Regression Analysis

TECHNICAL SKILLS

- **Programming**: Python (Pandas, NumPy, scikit-learn), SQL, R, C++, Java, MATLAB, SAS, Shell scripting
- **Framework & tools**: Git, Bash, Excel, SPSS, Docker, A/B Testing, Hadoop, HTML, CSS, JavaScript

INTERNSHIP EXPERIENCE

Harvest Fund Beijing, China

Research Assistant - AI Lab *April 2021- July 2021*

- Researched on artificial-intelligence-based financial product of other funds with competitive relationships
- Used Python and MySQL to calculate Piotroski F-Score and help with Hong Kong stock investment decision
- Applied K-Means and DBSCAN models to partition stocks into different clusters, further increased the interpretability of the team's investment products
- Employed quadratic programming in Python to optimize funds' position by minimizing the L_2 norm between net asset value and stocks' market value, achieved 2% mean absolute error

State Information Center Beijing, China

Research Assistant - Postdoctoral Workstation *July 2020-March 2021*

- Re-implemented Machine Learning algorithms in literature related to international investment matching
- Used multi-threading programming in Python to quickly collect unstructured research data from UN' website
- Built data visualization pipelines with Python (Pyecharts, Pyplot) displaying models' results automatically

PROJECT EXPERIENCES

Stock Price Prediction Based on Natural Language Processing Published on *Complexity* (Link)

Advised by Prof. Xiaobin Tang *November 2020-November 2021*

- Proposed innovative finetune strategy to generate keywords with better interpretability and predictability of CSI 300 Stock price from original seed words' Wikipedia
- Finetuned BERT and NEZHA in Python (TensorFlow) based on more than 500,000 sentences in the training set from Chinese GLUE and obtained 90.06% accuracy on over 19,000 sentences in the test set
- Used Pearson correlation coefficient and lagging terms to remove noises in data and select predictive variables
- Enhanced LSTM prediction performance with 28.20% RMSE decrease via new keywords' Google Trends

HONORS & AWARDS

- National Scholarship, China (Top 1%, 2020)
- University Outstanding Student, University of International Business and Economics (Top 5%, 2020)
- First Prize, Contemporary Undergraduate Mathematical Contest in Modeling (Top 1%, 2020)
- First Prize, The Asia and Pacific Mathematical Contest in Modeling (Top 5%, 2019)
- First Prize, The 11th Chinese Mathematics Competitions (Top 7%, 2019)

ACTIVITIES

- **Chairperson**: Quant Factory Student Investment Club, University of International Business and Economics

Yutong (Oliver) Xu
(952) 212-9255 | oliver.yu.xu@gmail.com | linkedin.com/in/olixu | github.com/lnconsistent

Education

New York University | New York, NY *September 2022 – May 2024*
M.S. in Data Science

New York University | New York, NY *September 2018 – May 2022*
B.A. in Mathematics GPA: 3.5
Minor in Computer Science

Experience

Center for Neural Science at New York University | NY *December 2019 – May 2022*
Research Assistant
- Collaborated on project investigating retinal flow in natural locomotion
- Implemented computer vision algorithms in Python and MATLAB to process and analyze field data
- Developed MATLAB simulations of natural locomotion to generate auxiliary data for hypothesis testing
- Created MATLAB and Python scripts to visualize data and results

Maxvision Technology Corp. | Wuhan, China *July 2019 – August 2019*
Computer Vision Research Intern
- Researched state-of-the-art methods in artificial intelligence for 3D face recognition and verification
- Processed image databases in Python to train and test machine learning and computer vision algorithms

Garcia Research Program at Stony Brook University | NY *June 2015 – September 2017*
Student Researcher
- Collaborated on project investigating the cytotoxic effects of titanium dioxide nanoparticles; found evidence suggesting nanoparticles did have adverse effects on human cells
- Recorded and analyzed microscopy data; submitted paper to the Siemens Competition and placed as a semifinalist

Projects

Heading perception and the structure of the optic acceleration *May 2022*
Co-authored paper investigating the use of the singularity of the optic acceleration field to estimate heading in a simulated subject. The preprint of the paper is available at doi.org/10.48550/arXiv.2204.12607.

Distance Estimation from RGB Images *October 2019*
Built a deep neural network in Python that generates an estimate for the distance between the camera and specific landmarks in images. Estimates generated by the network correlated directly with the real-world displacement of identified objects. Code available at github.com/lnconsistent/cv-object-distance-estimation.

Skills

Languages MATLAB, Python, Java, C++

Frameworks/Libraries NumPy, OpenCV, PyTorch, SciPy, and others

Achievements and Awards

NYU Tandon School of Engineering Dean's List *2018 – 2019*
National Merit Finalist *2017*
National AP Scholar *2015*
Semifinalist at the Siemens Competition *2015*

Palak Bansal

Jersey City, New Jersey · 5513259195 · palakb1406@gmail.com · LinkedIn

Professional Summary - I am experienced in Java and Python and have extensive knowledge of software development along with machine learning and deep learning libraries. My communication skills are excellent and I have spoken about Data Science at multiple conferences. I am looking forward to exploring different problems in new industries to support my expertise while I complete my MS in Data Science at NYU.

Education

New York University, New York
Master of Science, Data Science ***2022-2024***
Current coursework: Optimization and Computational Linear Algebra, Intro to Data Science, Probability and Statistics
Expected coursework: Machine Learning, Big Data, Database Management, Natural Language Processing

Manipal Institute Of Technology Manipal, India
B.Tech. Information Technology, CGPA: 8.46/10 ***2015-2019***
Minor specialization in Data Analytics
Relevant coursework: Machine learning, Computer vision, Pattern recognition, Information Retrieval, Business Intelligence Big data analytics
Additional coursework: 'Neural Networks and Deep Learning' by DeepLearning.AI on Coursera
Best student award: President of student club ISTE, Manipal

Skills

Languages : C++(Advanced), Java(Advanced), Python(Intermediate)
Technical Skills : NLP, Deep Learning, SQL, Graphql, Microservices, Data Structures, Algorithms, Tensorflow, Pandas, Numpy

Experience

SAP Ariba Bangalore, India
Developer Associate ***2019-2022***

- Contributed as **backend developer to a spring boot application** that performs optimization and analysis of procurement data.
- Generated real-time intelligent insights for customers using **unsupervised learning, anomaly detection and feature selection**.
- Designed and developed multiple business critical features, **GraphQL and REST APIs** for exporting data, enabling note sharing, connecting to multiple data sources, integrating with other microservices and multivariate visualization.

Data Science Intern ***2019***

- Performed **NLP** tasks including **Named Entity Recognition**, **Text Classification**, **Similarity Calculation** to extract information by **transforming unstructured and differently structured data** to the structure required by the SAP system.
- Researched and applied various NLP tools and techniques including **Pandas, NLTK, GloVe, BERT, ElMo and IBM Watson Knowledge Studio** to perform similarity calculations for column mapping with **82% accuracy** and **NER** to identify entities from unstructured customer data with **0.76 precision and 0.62 recall**.
- **Won the poster presentation** for this at **WE Local India** conference organized by SWE, Society for Women Engineers.

Reliance Jio Mumbai, India
Data Science Intern ***2018***

- **Preprocessed, analyzed and visualized data** logs generated by Netra AAT(Automatic Antenna Tuner) using **R and Python** working on libraries like **Tensorflow, Pytorch and Matplotlib**.
- **Applied neural networks** using **Keras** to classify Netra's reading to **predict reliability of results** and improved the performance by **20%, rejecting the less reliable readings.**

Engineering Projects

Intelligent Insights Generation ***2021***

- **Built a Django application** to generate insights from customer data which provides information about the distribution, highlighting exceptional instances and important features **in real time with unsupervised learning**.
- Researched and implemented various **outlier detection algorithms** and a **feature selection algorithm** specific to the use case.
- Presented the work in **SAP's largest conference SAP DevX** as a part of the Predictive Summit.

Information Extraction from unstructured data *2017*

- Developed a tool to **perform topic modeling** to identify topics from academic corpuses in sample texts using **LDA and Mallet.**
- Implemented and compared all the existing word similarities and sentence similarities based on **NLTK WordNet**.

Price Prediction *2019*

- Predicted prices of various items grouped in categories within a $5 margin using times series data spanning multiple years.

Achievements and Awards

- **Society of Women Engineers - WE Local Bengaluru conference** *2022*

  Panelist and primary speaker for "Communication skills for women in tech" at the SWE (Society of Women Engineers) WE Local Bengaluru conference 2022.

- **Global Codeathon And Ideathon At Sap Ariba** *2020*

  Made it to the finals for creating a prototype of an intelligent tool generating insights and performing intelligent supplier recommendation.

- **Coding Contest Winner** *2019*

  Won a coding contest focusing on competitive coding organized by Honeywell at the WE Local India conference in Bengaluru

---

Extracurriculars and Community service

- **"Unconscious bias" Seminar at SAP-** Delivered a seminar on "Unconscious bias" at my workplace with over 150 attendees. *2021*
- **Empower Project -** Organized an initiative for women in technology to improve their communication skills. Mentored the first batch of 10 women through well structured sessions involving technical discussions and blog writing. *2020-2021*
- **Global Youth Leadership & Girl-child Foundation (GYLGF) -** Volunteered to mentor young girls in Nigeria through the Raising Girls Ambition (RAGA) project by Global Youth Leadership & Girl-child Foundation (GYLGF) *2021*
- **Technovation Idea Labs, by UNESCO, TECHNOVATION and SAP -** Volunteered for Technovation Idea Labs – AI/Entrepreneurship program to empower girls with artificial intelligence (AI) entrepreneurship skills. *2020*
- **Rotaract Club -** Volunteered for clothes distribution drives, orphanage and old age home visits at Rotaract Club, Manipal. *2015-2017*

POOJA ARYAMANE

New York, NY 10009 | paa9751@nyu.edu | +1 (351) 220-8041 | GitHub: pooja-aryamane

EDUCATION

MS in Data Science, May 2024
New York University, Centre of Data Science, New York, NY
Relevant Coursework: Introduction to Data Science, Programming for Data Science, Optimization and Computational Linear Algebra

MS in Applied Statistics and Data Analytics, June 2022 **(CGPA: 3.86/4.0)**
NMIMS University, Mumbai, India
Relevant Coursework: Generalized Linear Modelling, Statistical Computing in Python and R, Data Mining, Machine Learning

BS. Applied Statistics and Data Analytics, June 2020 **(CGPA: 3.74/4.0)**
NMIMS University, Mumbai, India
Relevant Coursework: Probability Theory, Discrete Mathematics, Multivariate and Univariate Calculus

TECHNICAL SKILLS

Python, R, Base SAS, C++, Swift, HTML | SQL, SPSS | Tableau, SAS VA, Scilab | Machine Learning, Deep Learning

PROFESSIONAL EXPERIENCE

New York University | *Course Assistant; Regression II Categorical Data Analysis* **September 2022 – December 2022**
Enabled students to understand the course material and facilitated the preparation of new content.

Indian Institute of Technology Bombay (IITB) | *Machine Learning Researcher* **March 2021 – July 2022**
- Collaborated with the Machine Learning team to develop a tool that automates the digitalization of documents using computer vision and deep learning.
- Achieved an overall accuracy of **89%** for OCR, and an mAP of **91%** for document layout detection
- Built and evaluated state-of-the-art deep learning architectures using **Tensorflow, Pytorch, and Detectron2 in Python**.

Sutherland Global, Mumbai, India | *Analytics Division Intern* **May 2019 – July 2019 | July 2020 – August 2020**
- Collaborated with the analytics team to implement and deploy a predictive **logistic regression** model to identify fraudulent ticket sales for a major commercial airline.
- Devised a pipeline for data cleaning, feature building, and modelling using **R**, achieving an accuracy of **96%** on monthly ticket sales.
- Presented an outcome report, detailing detailed a potential monthly reduction in revenue loss of **25%** on model deployment.

SELECTED PROJECTS

Empty Space Detection for Inventory Management
- Optimized inventory management in retail stores by developing an interface designed to alert store managers on stock-outs, powered by deep learning and computer vision; achieving an mAP of 82%.
- Developed a robust object detection algorithm to detect empty spaces for restocking using the Faster-RCNN architecture.

Sign Language Recognition using Hidden Markov Models
- Built an algorithm to convert the hand and face movement during sign language (ASL) into English words to enhance communication between hearing and non-hearing individuals.
- Utilized Hidden Markov Models and deployed the algorithm in real-time.

Real-Estate Data Modelling Using Structural Equation Models
- Modelled **50** features that affect real-estate pricing in Mumbai, India by analysing their combined and individual impacts.
- Collaborated with a small team to analyse an extensive, 10GB dataset provided by JM Financial Real-Estate using SQL and developed a Structural Equation Model, predicting property value based on various factors with an accuracy of **92%**.

LEADERSHIP AND ACHEIEVEMENTS

Positions of Responsibility
- Co-Director of Events in the Graduate Community Student Building Group at New York University. [2022]
- Head of Marketing in the Student Council at NMIMS University. [2021-2022]
- Events Member of the Entrepreneurship Cell at NMIMS University [2020-2021]

Academic Achievements
- Winner of a 24-hour hackathon, DigiHack 202, organized in NMIMS University.
- Completed AI for Medical Diagnosis Online – DeepLearning.AI 2020

Evan (Yuheng) Yang

Email: yy2597@nyu.edu | **Mobile:** +1 (201) 748-9638 | **LinkedIn:** www.linkedin.com/in/evan-yang-a8825a1a7

EDUCATION

New York University New York City, NY

M.S. in Data Science | **Cumulative GPA: 4.0** September 2022 – May 2024 (Expected)

- Coursework: NLP with Representation Learning, Optimization, …

New York University Shanghai Shanghai, China

B.S. in Data Science; B.S. in Business and Finance | **Cumulative GPA: 3.927** September 2018 – May 2022

- Awards: **Summa Cum Laude (Top 5% ranking)**, University Honors Scholar, NYU Shanghai Global Quintessence Scholarship
- Coursework: Probability and Statistics, Databases, Machine Learning, Econometrics, Investments, Derivatives, Corporate Finance, …

WORK EXPERIENCE

Ziroom | *the largest unicorn in IoT & housing management in China* Beijing, China

Data Scientist Intern, AI and Business Intelligence Dept. June-August 2021

- **Data modeling with machine learning:** Worked on an inventory turnover diagnosis project, developed a well-performing classification model based on **Random Forest** and **Neural Networks** to predict whether the inventory unit can be leased with a test ROC-AUC of 0.8. The model was integrated into operations applications and used by retailing teams with **over 1500 workers**.
- **Data analysis:** Analyzed and evaluated the factors affecting inventory turnover. Built a factor improvement advising model, and reported the analysis results to retailing teams for refinement and adjustments to inventory units. The prototype with the classification model helped improve overall inventory turnover performance by **20%**.
- **Data management and visualization:** Performed data retrieval and analytics using **Hive**, assisted a data warehouse architect with schema modeling. Completed visualization tasks using **Tableau** and assisted building 2 dashboards using **Youdata BI**.

Essence Securities | *Top 10 investment bank in China* Shanghai, China

Security Analyst Intern, Research Dept, June-August 2020

- **Industry and company research:** Participated in the research of the company POPMART, composed industry analysis reports (partially published).
- **Research report composing and data analysis:** Completed qualitative research on supply chains and logistics systems for Yonghui Superstores (SH: 601933) under advising. Assisted data processing and analytics for valuation modeling.

PROJECTS & RESEARCH

Kaggle *Super Soaker* Performance Prediction Challenge August 2022

- **Result:** Ranked **top 7%** (131/1888) solo.
- **ML pipeline implementation:** Obtained insights among different features, performed feature engineering, and applied an ensemble model using Logistic Regression, Random Forest, and Neural Networks.

Financial Distress Prediction based on a Time-weighted Ensemble Model and Its Application January-May 2022

- **Features Construction:** Preprocessed over 50 financial gauges of 490 companies across different years, applied feature engineering using PCA, and performed sentiment analysis to MD&A from the financial statements of these companies to construct features pool.
- **Model design:** To predict default risks of these companies, created a **time-weighted ensemble classifier** that dynamically assigns different importance to different base classifiers (logistic regression, Random Forest, etc.) with optimal feature choices across different years. The results show the proposed algorithm could predict the potential default three years after at a test ROC-AUC of 0.8693.

MBTI Personality Categorization Prediction with Social Media Speech Posts November-December 2021

- **Unstructured data processing:** Applied text preprocessing techniques using NLTK to unstructured forum posts by **over 8600** individuals for the four-dimensional classification tasks.
- **Model training and evaluation:** Trained and fine-tuned classifiers based on Random Forest and LSTM with the preprocessed data to complete the personality prediction. Achieved **test accuracy of 80-90%** in each of the four theoretical dimensions.

SKILLS

Programming and Toolkits: Python, Scikit-Learn, PyTorch, Numpy, Pandas, Stata, SQL, Basic AWS Architecture, JavaScript, D3

Language: Mandarin (Native), English (Fluent, TOEFL 112), Japanese (Beginner)

Chaitali Joshi

cpjoshi202@gmail.com | +1 (201) 936 7618
https://www.linkedin.com/in/chaitali-joshi-b25499169/
Jersey City, NJ, USA

EDUCATION

New York University, *New York City, NY* *Aug 2022-May 2024*

- **Master of Science in Data Science**: *Introduction to Data Science, Probability & Statistics, Optimization & Computational Linear Algebra*

Veermata Jijabai Technological Institute (VJTI), *Mumbai, India* *Aug 2017-May 2021*

- **Bachelor of Technology in Electrical Engineering**; Projects in **Machine Learnin**g and **Deep Learning**
  GPA: 8.47 / 10

WORK EXPERIENCE

Business Analyst, Axis Bank, *Mumbai, India* *Aug 2021-Jul 2022*

- **Created** and **maintained** the ETB Personal Loan XSell **Dashboard** on **SAS Viya** on a monthly basis **by performing data analysis** using **SAS** and **SQL** to **present key metrics** pertaining to business overview, conversion rates, base synthesis and risk of the ETB Personal Loan business to various business stakeholders
- Prepared the Retail Asset without Saving Account Pre Qualified customer database for Personal Loan by extracting,merging and filtering data from source tables on a monthly basis
- **Performed deep-dive analysis** of the ETB Personal Loan business risk **using SAS and SQL to identify the segments** that were driving the trends observed in the early risk data by **analyzing the risk split across different customer segments** and presented the findings to stakeholders

Research Assistant, Center of Excellence (CoE-CNDS), VJTI, *Mumbai, India* *Dec 2018-May 2021*

- **Developed an Autoencoder(AE) based semi-supervised learning model** in **Python**, to **detect fraud** and **anomalies** in credit card transaction data, that was **shown to have a lower RMSE as compared to traditional Machine Learning algorithms**.
- **Deployed LGBM model,** using **Python to forecast load demand** given various weather parameters as features; **performed feature engineering** to **create new features from existing features.**

Technology Consultant Trainee, PwC India,*Mumbai, India* *May 2020-Jul 2020*

- **Forecasted** the **sales** of various items at stores **using Machine Learning** and **Analytics capabilities** and **AI-guided tools** for **data analysis**, **visualization**, feature **engineering** and **model training within SAP Analytics cloud.**

PROJECTS

Audio and Video Deepfake Detection | *Python, Keras,Tensorflow,ML,DL,Nvidia DGX-1* *Aug 2020-May 2021*

- **Designed two systems** to accurately **differentiate** between **real** and **fake audio** and **video on Nvidia DGX-1**.
- **Implemented two methods** for detecting fake audio: **a feature-based approach utilizing ML algor**ithms and a**n image-based approach employing DL algorithms, notably TCN, which gave a test accuracy of 92%.**
- **Applied Transfer Learning technique to extract features** from video data and **implemented LSTM and TCN models** to detect fake video.

Publication **-** A Deep Learning Framework for Audio Deepfake Detection, Springer Journal

DeepWind: Wind speed forecasting | *Python, Keras, Feature Engineering,ML* *Jan 2020-April 2020*

- Obtained wind speed forecasts for Indian weather stations **using ensemble learning on ML algorithms**: the **LGBM** and **LSTM networks**,**employed Miss Forest, used unique feature representation** to emphasize recent data and FFT with **Digital Filters to remove outliers in the data.**
- **Presented the project to a panel of 5 industry experts,** at the "CDAC - NVIDIA AI Hackathon", **ranked top 10 across India among 350+ teams.**

Publication **-** A Short-term Wind Forecasting Framework using Ensemble Learning for Indian Weather Stations, IEEE INOCON

Explainable AI Project |Python, *Tensorflow, Keras, Explainable AI* *Oct 2019-Dec 2019*

- **Employed Explainable AI techniques like LIME & LRP** to **interpret AE** trained on a gas pipeline system to identify attack scenarios.

Publication **-** Interpreting a Black-Box Model used for SCADA Attack detection in Gas Pipelines Control System, IEEE INDICON.

Anomaly Detection in Gas Pipeline Systems | *Python, Tensorflow, ML, DL* *Jun 2019-Aug 2019*

- **Built an autoencoder network as a DL approach to anomaly-based** Intrusion Detection Systems to **detect attacks** in SCADA data that is used to control gas pipeline systems that o**utperformed traditional ML algorithms.**

Publication **-** A Semi-Supervised Approach for Detection of SCADA Attacks in Gas Pipeline Control Systems," IEEE-HYDCON

SKILLS

Languages: Python, C++, SQL, SAS ***Others***: Machine Learning, Deep Learning, Computer Vision, Nvidia-DGX
Frameworks/Libraries: Keras, Pandas, Scikit-Learn, Matplotlib, Numpy, Tensorflow, Excel, MS Office, SAS Viya

Shreemayi Sonti

ss16270@nyu.edu | 551-998-3256 | https://www.linkedin.com/in/shreemayi-sonti-076617157/

EDUCATION

New York University **September 2022 - May 2024**
MS in Data Science
Coursework:
Introduction to Data Science | Computational Linear Algebra and Optimization | Probability and Statistics
Machine Learning | Big Data

Jawaharlal Nehru Technological University **July 2017 - July 2021**
Bachelor of Technology- Electronics and Communication
GPA 9.3/10
Coursework:
Calculus I & II | Linear Algebra | Transform Calculus | Embedded Systems | Digital Communication | Digital Signal Processing | Intro to IoT | Algorithms and Programming using C | Data Structures using Python | OOPS using Java | Computer Networks | Principles of Cloud Computing

SKILLS

Languages: Python, PL/SQL, Java | **Data Visualization and Analytics:** PowerBI | **Cloud:** Azure- Blob Storage, Databases, Key vault, VMs, AWS - lambda | **Databases:** MS SQL, Oracle | **Version Control:** Git (GitHub) | **CI/CD Build Tools:** Jenkins, Gradle | **Other:** JIRA, MS Office

EXPERIENCE

Gap Inc., Hyderabad **July 2021 - July 2022**
Software Engineer
Worked on internal applications for Store Associates
- Generated reports from huge volumes of sales and inventory data using data engineering and visualization tools like Apache kafka, SQL, and Power BI. Shared reports with leadership to support inventory allocation and sales analysis.
- Led the migration of the application database from Oracle Cloud to MS SQL on Azure, modified the procedures, code, CI/CD pipeline accordingly.
- Updated the ARM template for storage accounts on Azure and added expiration rule to purge files older than certain days automatically eliminating the need for deleting files manually or paying for the extra storage costs.

Collins Aerospace, Bengaluru **April - July 2021**
Intern- Data Analytics team
- Developed a decoder logic for analytics team using Python that encoded sortie data and deployed it on AWS lambda.
- Compared Kafka and RabbitMq as containers on Docker and decided Kafka is a better replacement to the existing system. The existing system received files in a VM and processed the files to get the required data.

PROJECTS

Comparison of Statistical and Deep Learning Techniques in time series forecasting **July 2021**
Modeled using two different types of time series algorithms and compared the results- S/ARIMAX and LSTM on daily (stock price) and monthly (champagne sales) datasets and observed that LSTM outperformed incase of stock price prediction, whereas S/ARIMAX did better with monthly datasets.

IoT based Home Safety and Security **Jan-March 2020**
Periodically checking if the window is open and if there is any gas leakage in the house and alerting the user using a mobile app (Kodular companion). Data read by the sensors is pushed to Google firebase, a real-time database using LoRa communication technology.

Smart health monitoring system Hackathon **Jan 2019**
The project deals with monitoring parameters of health using Raspberry Pi by taking data via sensors and hence, suggesting remedies for better health.

CERTIFICATIONS

- Certified Associate in python programming (PCAP)
- AWS academy Cloud foundations course
- Python for Data Science and Machine Learning on Udemy by Jose Portilla
- Introduction to IOT by NPTEL, Duration- 12 weeks July-Oct 2019 and received a gold medal from IIT KHARAGPUR
- Introduction to Networks course on Cisco Netacad.

Ankush Chaudhari

+1 201-238-7751 | GitHub | 58 Tuers Ave. Apt 2
ac9820@nyu.edu | LinkedIn | Jersey City, NJ-07306

EDUCATION

New York University - Center for Data Science — New York, USA
Master of Science, Data Science - Industry Concentration — Aug 2022 – May 2024
Coursework: Optimization and Computational Linear Algebra, Probability and Statistics, Intro to DS

Vellore Institute of Technology — Vellore, India
Bachelor of Technology, Computer Science and Engineering | GPA: 3.97/4.00 — 2017 - 2021
Coursework: Data Structures & Algorithms, DBMS, Machine Learning, Image Processing, NLP, Web Mining

EXPERIENCE

Snackpass | *Software Engineer – Data Squad* — Sep 2021 – May 2022 | San Francisco, CA (Remote)

- Built optimized **ETL pipelines** using **Airflow** on **AWS** to integrate 5 different data sources **handling 50M+ records/day**.
- Designed and built a **Leads Scoring internal tool** which uses raw data points of over **1.5M restaurants** to compute the score of their probability of conversion to Snackpass partners based on features such as orders pattern, online reviews across 18 sources, menu, etc., **saving the sales teams' approx. 3000 worker-hours/year.**

Sleek | *Machine Learning Engineer* — Aug 2020 – Sep 2021 | San Francisco, CA (Remote)

- **Spearheaded R&D of patent-pending technology** which generates crowd analytics for large areas by using ML to analyze radio wave patterns emitted by smartphones with up to **85% accuracy** deployed on cost-effective portable hardware.
- Developed a **CNN-based client menu digitization** solution which **reduced onboarding time from 2 hrs. to 10 min.**
- Developed an **XGBoost regression-based order preparation time estimator** model trained on features such as concurrent open orders count and items specifications, current order items specifications, historical trends of the store, etc. to estimate the time duration required to complete the current order. **Achieved an RMSE of under 2 min.**

Sleek | *Data Scientist Intern* — May 2020 – July 2020 | San Francisco, CA (Remote)

- Developed a **CNN-based crowd analysis application** to detect queues and estimate wait times **with an RMSE of 2 min.**
- Optimized deep learning processes to support **GPU-based inference**, which **increased the throughput by 10X.**

VIT in collaboration with **ISRO | *Student Researcher*** — Jan 2021 – July 2021 | Vellore, India

- Worked under the guidance of Dr. Santhi V. on the **SAR Satellite Image Analysis** research project**,** proposed and sponsored by the **Indian Space Research Organization** (ISRO).
- Developed the **sidelobe suppression algorithm** for lossless enhancement of 16-bit satellite images using Python.
- Formulated a novel approach that significantly improved the quality of results and **reduced the processing time from 8 hrs. to 15 min. for gigapixel resolution SAR images** compared to the existing solution.

GSPANN Technologies Inc. | *Data Scientist Intern* — May 2019 – Jun 2019 | Hyderabad, India

- **Developed an image fashion product search MVP for Kohl's** by training a CNN to extract image features and cross-referencing them with the products database which **reduced user search time by 40% compared to text-based search.**

PROJECTS

FITBOT – AI-powered personal fitness coach — Mar 2021 – Jun 2021 | India

- Solves the issues with fitness centers especially in the COVID-19 situation while maintaining users' safety and privacy.
- Generates personalized visual guidance on display to train users to perform the exercise correctly. [Demo Link]
- Used a pre-trained model (OpenPose) to detect the user's posture and developed an algorithm to evaluate its correctness.

Social Distancing Analyzer COVID-19 — Apr 2020 – Jun 2020 | India

- Developed an AI tool that monitors social distancing and mask-wearing protocols in public places using CCTV surveillance
- **Deployed in 3 hospitals** to assist and alert security personnel when required in real-time.
- **400+ stars and 200+ forks on GitHub** [Link] and also **featured in The Times of India.** [Link]

Automated Toll and Fine Collection System — Jan 2020 – Nov 2020 | India

- **Patent published** (not granted yet) solution which uses a combination of deep learning-based algorithms and infrastructure design to automate toll and speeding fines collection. [Patent Application No: 202041055717 A] [Link]

LEADERSHIP

Zetatron Solutions | *Co-Founder and CTO* — Sep 2019 – May 2020 | India

- Developed products that **helped 30+ SMBs to make data-driven strategies** to maximize revenue.
- Directed a team of 6 employees and **generated $15K+ revenue in 9 months.**

Society of Petroleum Engineers – VIT | *President* — Jan 2019 – Feb 2020 | India

- Directed a team of 35 members and established efficient workflows to achieve the organization's goals.
- For leading the chapter to **top 5% globally, received Presidential Award for Outstanding Student Chapter.**

SKILLS AND INTERESTS

Programming Languages: 5000+ Lines: Python, SQL | 1000+ Lines: Java, MATLAB
Libraries: OpenCV, NumPy, Pandas, Airflow, PyTorch, Scikit-learn, CUDA, BS4
Technologies/Areas: GCP (BigQuery, Compute Engine, Composer), AWS (MWAA, S3, EC2), Computer Vision

Mingxuan (Curly) Wu

Email: curlywu99225@gmail.com | Mobile: (+1)858-295-9303 | Website: https://www.linkedin.com/in/curly-wu

EDUCATION

New York University, Center for Data Science (CDS) Sep 2022 – Dec 2023
Candidate for the **M.S. in Data Science**

University of California, San Diego (UCSD), CA Sep 2018 – Mar 2022
B.S. in Data Science and **B.S. in Management Science**
Overall **GPA of 3.91 / 4.0**
Cum Laude Distinction, UCSD Provost Honor for 11 Quarters, Member of Phi Beta Kappa Sigma Chapter of California

WORK EXPERIENCES

Data Scientist Intern | Franklin Templeton Investment Jun 2022 – Aug 2022

- Extracted-transformed-loaded FRED data, **leveraged time series analysis** to forecast the Inflation rate, handled **post-pandemic data abnormality** and achieved a **45.6% improvement in MAE**:
  - Examined and transformed 50+ macroeconomic variables based on **ADF tests** scores, filtered out 30+ variables that linearly aligned with the target using regression P-values, and dropped highly correlated variables using **VIF tests**
  - Tracked data abnormality during the pandemic and aligned the data by capping outliers and feature selections
  - **Time Series Spited data** and **tuned hyperparameters** of models including **Lasso, Ridge, Random Forest, and Light GBM**, achieving a 45.6% improvement in MAE over the **empirical Baseline**
  - Visualizing modelling results and creating **workflow charts** for a data science presentation to business users

Data Analyst Intern | JDD Tech Company Jul 2021 – Sep 2021

- Assisted the collection and analysis of customer behaviors data**:**
  - Processed **high volumes of customer data** to capture user persona at JDD in support of digital marketing strategies
  - **Transformed Flume HDFS Interceptor to Flink HDFS Interceptor** to improve data throughput by 200%
  - Leveraged 10+ software, including **GitHub**, **Hive**, IDEA, **MySQL**, **Spark** Core, Flume, Flink, Navicat, Kafka, and Scala language to accomplish the tasks; performed batch job scheduling using Azkaban Hadoop

Data Analyst Intern | Alibaba Group Jun 2020 – Aug 2020

- Completed a data-driven comparative study on the topic of live commerce vs. traditional e-commerce in China:
  - Performed **data mining** and **analyzed 6+ forms of KPIs** such as page view (**PV**), unique visitor (**UV**), daily active user (**DAU**), monthly active users (**MAU**), peak concurrent users (**PCU**), average revenue per user (**ARPU**), etc.
  - Queried and organized large-scale customer data with the aid of SQLite relational database
  - Conducted customer review analysis and **sentiment analysis** via cross-validated **random forest, gradient boosting, logistic regression** with texts **TF-IDF encoded**, and Naïve Bayes models

PROJECTS

Team Leader | Blockchain and Smart Contract Application: Gym Coin
Data Science Senior Capstone Design, UCSD Sep 2021 – May 2022

- Developed a **Blockchain**, **Smart Contract**, and **NFT** application for commercialization of daily exercises:
  - Researched on cryptocurrency tokens (ERC20 and ERC721) and their decentralized applications, developed an innovative solution that decentralizes the exercise-reward systems, and synchronize the ideas into a whitepaper
  - **Developed** and **tested** the smart contract on **Remix** using **Solidity**, implemented the contract on Ethereum ropsten test network, and deployed and shipped on Scaffold-ETH for front-end User Interface: http://dsc180a03_gymcoin.surge.sh/

Team Leader | Business Analytics for GoShare Sep 2020 – Jan 2021

- Established a data-driven workflow to support the last mile logistics platform at **GoShare**:
  - **Enabled automated data mining** from the Metabase business intelligence tool at GoShare, focusing on evaluating project completion & cancellation rates and service ratings while considering geospatial and partnership data
  - Implemented **descriptive data analytics** and **visualization** on the collected data to drive performance evaluation for GoShare employees
  - Performed **diagnostic and predictive data analytics** to forecast the acceptance rates, incorporating the rate of return and credit scores of customers into the models to optimize resource allocation and pricing

Team Leader | Cloud/Cluster Computing Course Project, UCSD Sep 2020 – Dec 2020

- Built a **million-scale data-centric pipeline on AWS** to predict product ratings and **optimize marketing strategies**:
  - Processed more than **45 million** data points by cleansing and **flattening** the semi-structured data, **imputing** missing data, and conducting **PCA** to reduce dimensionality redundancy, and applying **one-hot encoding** to handle sentiments
  - Trained and validated supervised models such as **logistic regression**, **random forest**, and **decision tree** to drive predictions
  - Conducted hyperparameter tuning and cross-validation to **boost accuracy to 81.5%**

SKILLS

Programming Languages: Python (**NumPy**, **Pandas**, **Scikit-Learn**, **PySpark**, **Matplotlib**, Altair), **SQL, R** (ggplot, dplyr), Java, MATLAB, Stata, Scala, Dask.
Software: Jupyter Notebook, Git, IntelliJ, Microsoft Office, Visual Studio, PyCharm, Eclipse.

SAAHIL JAIN

New York, NY | (551) 344-6681 | sbj7913@nyu.edu | https://www.linkedin.com/in/saahiljain98

EDUCATION

New York University, GSAS, Center for Data Science, New York, NY — May 2024
Master of Science in Data Science
GPA: -

PES University, Bengaluru, India — May 2021
Bachelor of Technology in Computer Science and Engineering, Specialization - Data Science
GPA: 8.95 / 10
Scholarships: ***Prof. CNR Rao Merit Scholarship***

SKILLS

Coding Languages: Python, SQL, R, C, C++.
Operating Systems: Mac, Windows, Linux
Other Tools: TensorFlow, Keras, Scikit Learn, Numpy, Pandas, Google Workspace, Slack, Microsoft, Teams, Zoom.
Languages: Hindi (*Native*)

EXPERIENCE

Software Engineer, **Citrix,** India — 7/2021 – 8/2022
- Updated SSL protocols to ensure compatibility and better security for the Citrix ADC Service.
- Led a team of 6 employees to 1st place at the Citrix New-Hire Tech Project Competition.
- Designed a Slack chatbot equipped with natural language processing(NLP) to provide IT/HR solutions.

Software Engineer Intern, **Citrix,** India — 1/2021 – 6/2021
- Optimized the costs for Citrix by devising an architecture to create Citrix-managed ADC used by multiple clients to ensure maximum utilization.
- Validated AWS role ARNs of customers, to provide seamless ADC deployment and service.

Teaching Assistant (Machine Intelligence), **PES University,** India — 8/2020 – 12/2020
- Created video lectures explaining various topics in Machine Learning and Artificial Intelligence.
- Automated the assessment of assignments submitted by students.

Student Developer, **Microsoft,** India — 7/2020 – 8/2020
- Collaborated with 5 student developers from around the world on a project called Reimagining Healthcare, which personalized healthcare services in the post-covid world.
- Developed computer vision solutions to improve assistance for patients and ensure implementation of safety protocols.

Student Mentor, **PES Innovation Lab, PES University,** India — 5/2020 – 7/2020
- Counseled 3 student interns on a Neural Ordinary Differential Equations project.

Machine Learning Intern, **Microsoft Innovation Lab, PES University,** India — 6/2019 – 7/2019
- Developed an autonomous car on a simulator called Carla.
- Studied the effect of Reinforcement learning for the problem over the popular Imitation learning approach.

PROJECTS

Font Generation using Neural Cellular Automata — Fall 2020
- Generated new fonts for languages like Kannada having limited existing fonts.

Moving Target Interception using Multi-Agent Reinforcement Learning — Spring 2020
- Achieved coordination among autonomous agents to work together to catch a moving target programmed to avoid them.

Instrument Identification in Audio Clips using Machine Learning — Fall 2019
- Developed a software to identify 1 of 25 musical instruments from audio clips with various kinds of noise.

Hybrid Neural Network Library for Python — Spring 2019
- Designed and developed my own neural network to push the boundaries of regular neural networks.
- Developed a trainable neural network without the use of back propagation algorithm, by leveraging genetic algorithm.
- Achieved non-linearity in outputs without using activation functions by using an extra hybrid neuron in each layer with specialized functions to achieve the required results.

VOLUNTEER EXPERIENCES

- ***Community Volunteer***, Fly Higher India, Bengaluru, India — March 2019 – May 2022

Shikhar Rastogi

sr6644@nyu.edu | 2018843461 | |

EDUCATION

NEW YORK UNIVERSITY
M.S. in Data Science
Sep 2022 - May 2024 | New York, USA
Introduction to Data Science
Probability and Statistics for DS
Optimization and Computational Linear Algebra
Big Data (expected)
Machine Learning (expected)
Probabilistic Time Series Analysis (expected)
Natural Language Processing with Representation Learning (expected)

BITS PILANI
B.E.(Hons) in Computer Science
M.Sc.(Hons) in Economics
Aug 2015 - Aug 2020 | Goa, India
CGPA: 8.97 / 10.0
Neural Networks and Fuzzy Logic
Data Structures & Algorithms
Applied Econometrics

SKILLS

LANGUAGES
JAVA • Python • C++

TECHNOLOGIES
Spring Boot • MongoDB • Oracle SQL
PyTorch • Tensorflow • Pandas • Dask
PySpark • scikit-learn • Hive

CERTIFICATIONS

COURSERA
Deep Learning Specialization by deeplearning.ai
Natural Language Processing with Classification and Vector Spaces

EXTRA-CURRICULAR

Chief Coordinator at the Department of Publicity and Public Relations managing a 70 member team.

Trinity certified keyboardist and member of the Music Society of BITS Goa.

AWARDS

Excellent Performance Recognition Certificate at NTU

WORK EXPERIENCE

APPLE | Software Development Engineer
Jul 2020 – Jul 2022 | Hyderabad, India

- Deployed a model for multilingual hate speech detection to filter out hateful customer feedback, which achieved 90% accuracy in the real world using a fine tuned XLM-RoBERTa model
- Identified similar customer support queries using a fine tuned BERT based sentence transformer model to help advisors resolve customer issues quicker
- Developed resilient batch jobs using Spring Batch for the historical big data migration and update of close to 100 million records, with zero downtime reported
- Built fault-tolerant RESTful APIs using the Spring framework for systems handling close to 500 million service calls every day
- Won 2nd place in the Apple Worldwide Org Hackathon for developing a new product and coming up with unique business insights and metrics to add value

APPLE | Data Science Intern
Jul 2019 – Dec 2019 | Hyderabad, India

- Discussed the problem with business stakeholders to transform the business problem into an ML usecase and gathered requirements
- Performed extreme multi-label classification using Random Forests to populate fields in an internal application and reduced time taken in the flow by 20%
- Presented my work to the CIO of Apple and other VP level staff

RESEARCH EXPERIENCE

MULTIMODAL TURN TAKING SYSTEM FOR SOCIAL ROBOTS
Jan 2020 – Jun 2020 | Nanyang Technological University, Singapore
Under Prof. Nadia Thalmann via India-Connect at NTU Research Programme

- Used multimodal data to build an online continuous model to improve the flow in a conversation of a humanoid social robot
- Fused the prosodic speech and text features and fed them into a Bi-LSTM model with a custom attention mechanism to predict when the robot should speak up or interject

PROJECTS

ALPHABET RECOGNITION USING WIFI SIGNALS
Jan 2019 – May 2019 | BITS Pilani, Goa Campus

- Extracted data from interference in WiFi signals, and denoised and filtered using the 'signal' package on MATLAB
- Built a Siamese Network for One-Shot Learning with two identical convolutional neural networks to classify into different letters of the alphabet drawn in the air

MULTI-AGENT REINFORCEMENT LEARNING TO DEMONSTRATE COOPERATIVE HUNTING SCENARIOS IN ANIMALS
Aug 2018 – Dec 2018 | BITS Pilani, Goa Campus

- Set up a grid-world environment with multiple agents, with their respective actions and payoffs, and simulated the game using Multi-Agent Reinforcement Learning
- Trained the hunters using the Nash Q-Learning Reinforcement Learning algorithm and investigated when they cooperated or defected

Resume | Shivam Ahuja | MS Data Science

Center of Data Science, New York University | +1-3472071842 | sa7445@nyu.edu

PROFESSIONAL EXPERIENCE

ZS ASSOCIATES **New Delhi, India**
Decision Analytics Consultant **January 2021-June 2022**

- Estimated the demand landscape for a F&B client's portfolio in US and helped uncover $357M worth of vulnerable spaces utilizing "Jobs theory"; a new product has already been introduced in one space based off our suggestion
- Developed a classification model with 75% accuracy using a combination of MARS and Regularized logistic regression to predict the segments likely to prescribe a pharmaceutical vaccine which will be used by the sales field team for targeting
- Leading an internal team transition from a R based environment to a python one, updating the entire existing code-set

Decision Analytics Associate Consultant **July 2018-December 2020**

- Awarded "Upskill Guru" recognition from the ZS leadership for innovating a methodology called "Adaptive typing tool" which allows sales reps to identify HCPs segment by asking ~50% less questions compared to ZS's traditional approach
- Developed a messaging focused attitudinal segmentation using a decision tree algorithm for a mental health focused client and then overlaid it onto their target list by creating a logistic regression based prediction algorithm with 90%+ accuracy
- Developed, in a team of 4, the go-to-market strategy of an oral drug with expected peak sales of $4B by identifying key ZIP codes across U.S. using a k-means clustering algorithm for aggressive, post-launch targeting
- Identified the target demographic for a video calling device using a latent class clustering algorithm, leading a team of 3 to recommend the key opportunities and messaging, resulting in 3 more projects with the client and $750,000 in revenue
- Led a team of 5 to improve the messaging of physicians for a pharmaceutical company's portfolio of 3 insulin products by utilizing a combination of random forest, logistic regression and clustering algorithms, leading to a 20% increase in likelihood of endocrinologists prescribing one of the products, increasing its revenue by ~190%
- Implemented structural equation modelling in R programming language, which resulted in a 10% increase in share for the most receptive segment. Featured as the "Innovation of the Year" on ZS segmentation team's annual newsletter
- Coached and mentored a team of 19 associates across offices and conducted firm-wide training sessions on Excel and R

Decision Analytics Associate **June 2016-June 2018**

- Created an Excel-based model to identify the optimal sales force size for a printer manufacturing company, recommending an increase in staff, resulting in 10% more penetration for key pilot accounts
- Designed, in a team of 12, an unbranded direct-to-consumer (DTC) campaign for a medical product services company to increase the underlying condition's unaided awareness from ~40% to 68% amongst the affected population within 2 years
- Recommended targeting community oncologists over academic oncologists to a pharmaceutical company, further segmented this group using a decision tree algorithm, resulting in a 4% increase in share over the next year
- Identified consumer segments across 7 countries for portfolio of devices of a Fortune 500 software company using a latent class clustering algorithm, which resulted in 5 more projects and close to $3 million in revenue to the firm
- Delivered a visual dashboard using Tableau to report key survey metrics and KPIs, used for better customer targeting

VOLUNTEER EXPERIENCE

NeeV – A Learning Center (Part of ZS CSR activity) **New Delhi, India**
Volunteer **August 2017-December 2020**

- Organize monthly learning sessions on Mathematics along with a fun activity for children, aged 8-18, at a nearby school

LE ZS FC **New Delhi,**
Co-Admin **November 2016-June 2022**

- Started a weekly soccer group at ZS with 3 fellow players and have grown to 147 participants in five years

EDUCATION

New York University **New York, USA**
Master of Science, Data Science **2022-Present**

- Coursework: Intro and Programming in data science, Linear algebra, Optimization and Probability in data science

Thapar University **Patiala, India**
Bachelor of Engineering, Electronic and Communication, GPA 9.4 of 10 (3.76 of 4.0) **2012-2016**

- Awarded merit-based scholarship for being in the top 3 of 180 students in my branch
- Top 3 performers out of 30 interns for the 2015 HEP Training program at Mentor Graphics

ADDITIONAL INFORMATION

- **Certifications**: Completed "Python for Everybody" online course from Coursera and "Introduction to Football analytics" & "Modern Scouting and Data-Driven Recruitment" online courses from Statsbomb Academy
- **Interests:** Set game strategy for my local soccer team to reach two tournament finals; Finished amongst the top 5% of players worldwide in Fantasy Premier League in 2020-21; Expert at cooking/baking Indian desserts
- **Tools and Algorithms:** Proficient in R, Python, Tableau, Alteryx and Angoss Knowledge Seeker

SHULIN JI

+1 (201) 668-1150 | sj4016@nyu.edu | LinkedIn

EDUCATION

New York University, Center for Data Science New York, NY

Master of Science in Data Science Sept. 2022 – May. 2024 (Expected)

- Courses: Machine Learning, Big Data, Deep Learning, Fundamental Algorithms, Time Series, Causal Inference

Shanghai Jiao Tong University, School of Mathematical Sciences Shanghai, China

Bachelor of Science in Mathematics & Applied Mathematics (Zhiyuan Honors Program) Sept. 2018 – Jun. 2022

- Honors: Outstanding Graduate (top 3%); Academic and Leadership Merit Scholarship ($5000 awarded for the top 5%)

SKILLS

- **Programming and Software:** Python (scikit-learn, numpy, pandas, matplotlib, etc.), C++, MATLAB, SQL, Tableau, LaTeX, SPSS, Weka, Origin, Wind, Microsoft Office (Excel, Word, PowerPoint, etc.), Linux operations
- **Data Science Methods:** A/B Testing, Inference, Machine Learning, DS pipeline in business (cleansing, wrangling, visualization, modeling, interpretation), Statistics, Stochastic Process, Dynamic Systems and Numerical Simulation

PROFESSIONAL EXPERIENCE

PayPal Shanghai, China

Data Analyst Intern in Compliance Platforms Sept. 2021 – Apr. 2022

- Constructed self-inquiry **database dashboards** and official reports to **automate tracking functionality** on business volume and solution efficacy with **SQL** and **Tableau**
- Identified 10+ effective **live issues** to alert potential **algorithmic flaws** with detailed data detection and **interactive data visualization**, and guided further java-based modifications with engineers
- Raised 4 applicative **metrics** for classifying algorithms on imbalanced data with F-beta score, ROC and PR curve
- Designed an automatic **data report email system** for team engineers, and a detailed Compliance **data analysis illustration platform** to enhance communication efficiency between departments

Yinzhi Tech Shanghai, China

Data Engineer Intern in Product Department Jul. 2021 – Sept.2021

- Developed an intelligent platform to automate real-time analysis for investment consulting, including **data dictionary** construction, knowledge graph analysis and **interface design** in **Mockplus**
- Researched the investment diversity of fund managers: constructed a **web-crawler system** in **Python** for data collection, conducted label definition and further **regression analysis**

Lingjun Investment Shanghai, China

Quantitative Researcher Intern in Futures Department Apr. 2021 – Jul. 2021

- Generated 2 effective Supply/Demand **alpha factors** on commodities (methanol, rebar, etc.) in **Python** in **Linux**
- Formed a **data pre-processing** program with high generalization ability, later applied to all 48 futures commodities, **wrangling** all features in structure from unorganized details (diverse update frequency, delayed period, etc.)
- Customized **feature selection** with **product-based** knowledge, nearly tripling the raw volume, based on the former self-conducted Industry Analysis Reports for 48 commodities in Quantitative Futures Fundamentals

PROJECTS EXPERIENCE

Quantile Factor Model and Applications in Financial Technology Jul. 2021 – Jun. 2022

- Developed **Quantile Regression** on **Factor Model** with high-dimensional applications in **MATLAB**
- Generalized the 2-D model to fit 3-D scenarios, and captured quantile-dependent factors (unobserved by PCA) on empirical cases with heavy-tailed errors and outliers
- Implemented Quantile Factor Model on individual **stock prediction** (return rate) with 10 quantitative factor indicators (volatility, momentum, etc.), outperforming traditional Approximate Factor Model and PCA by lowering 8~9% MSE

Statistical Study on Typical Indicators of Credit Card Holders Mar. 2021 – Jun. 2021

- Launched statistical research on 24 features with **normality test, correlation test and non-parametric test**
- Applied Machine Learning methods (**LR, SVM, NN, RF**, etc.) for default probability prediction in **Python**, with further statistical analysis in precision, recall and F-1 score, reaching general accuracy at over 80%

Tayyibah Khanam

+1 (347) 556-6371 {New York, United States}

Email : tk2981@nyu.edu
Linkedin, Github
Personal Website

Education

- **New York University. Center for Data Science** — New York, NY
  Master of Science in Data Science — *Sep'22- May'24*
  Courses: Optimization & Computational Linear Algebra, Probability & Statistics for Data science, Introduction to Data Science
- **Aligarh Muslim University** — Aligarh, India
  Bachelor of Technology in Electrical Engineering; GPA: 9.07/10.0 — *Aug'18 - May 22*

Skills Summary

- **Languages**: Python, SQL, C, MATLAB
- **Tools**: Git, IBM SPSS, LaTex, Google Colab, Microsoft Azure, Jupyter Notebook, VSCode, Proteus, Simulink, macOS, Windows
- **Libraries**: Tensorflow, Keras, Pytorch, Scipy, Scikit-Learn, Matplotlib, OpenCV, Seaborn, Numpy, Pandas

Experiences

- **NYU Langone Health, New York University** — New York, NY
  Research Assistant @Fenyo lab, Supervisor - Prof David Fenyo — *Sept 22 - Present*
  - **Detecting Neutrophils**: Using Deep Learning Methods to Identify Neutrophil Infiltration in Histology Images to Helping Pathologists Diagnose Infection After Hip and Knee Transplant Surgery.
- **Algo University, Robotics Research Institute IIIT Hyderabad** — Remote
  Startup: Computer vision Intern — *Jan - July'22*
  - **Research**: Studied pose graph optimization, real-time molecular scene understanding (E-Net, U-Net) & depth prediction, with respect to the problem of bird's eye view multi-object SLAM for self-driving cars.
- **New York University, Center for Data Science** — Remote
  Summer Research Intern, Supervisor - Dr Elena Sizikova — *June - Dec'21*
  - **Visual ML**: Investigated interpretable ML algorithms, with a focus on special techniques for Visual Interpretability using GradCAM, NMF, GBP, etc. Further, evaluated and compared these methods using AOPC curves & BAM metrics.
  - **Medical Imaging**: Researched and reproduced results of popular reconstruction based methods (GANs & AEs) for anomaly detection in medical imaging.
- **Jawaharlal Nehru Medical College, Aligarh Muslim University** — Aligarh
  Summer Intern, Supervisor - Dr Ali Jafar Abedi — *June - Oct'21*
  - **Research**: Utilized, and compared results of numerous ML algorithms for feature learning and prediction of malnutrition in Indian under-five children on the Demographic Health Surveys dataset.
  - **Model development**: Introduced a novel anthropometric index **Stunted-Wasting** for prediction based on literature evidence along with the standard independent Stunting and Wasting indices. Further, achieved an Accuracy range from 74% up til **96%** and an AUC-ROC score range from 82% up til **99%** on the novel index.
- **Centre of Advanced Research in Electrified Transportation, AMU** — Aligarh
  Summer Research Intern, Supervisors - Prof Saad Alam — *June - Sept'20*
  - **Research**: Investigated the causes responsible for less popularity of EVs on road in Lesser Developed Countries through literature surveys. Further, performed a case study to analyze and gain insights on user charging behavior, electricity demand, and the effect of COVID-19 pandemic on both given a city's charging pile network.
  - **Optimization**: Developed a regression-based framework with a time-minimization approach that could optimize the EV Charging network of Mumbai by **18-26%**.

Publications

T.Khanam *et al.* "Efficient Machine Learning for Malnutrition Prediction Among Under-Five Children in India". Link

D.Mishra, **T.Khanam** and I.Kaushik. "Experimentally proven Bilateral Blur for Optimal Convergence". Link

T.Khanam *et al.*"Optimizing Electric Vehicle Charging With Charging Data Analytics". Link

T.Khanam *et al.* Big Data Applications in Smart & Sustainable Energy Systems: Review & Case Study. Link

Selected Project(s)

- **Fire Mapping using Satellite imagery (EY Data Science Challenge)**: Data centric approach to map fire regions on linescan datasets that involved extensive data pre-processing, clustering and masking on highly imbalanced GeoDataFrame datasets, aerial linescans and satellite imagery from NASA. Employed a custom made deep U-Net architecture built on dice coefficient loss for the binary segmentation task. Achieved an accuracy of **74%** & qualified to finals with a **world rank = 20**. (May'21)

Awards & Leadership

- Selected in top 25 (out of 330) applicants of university for the Sir Syed Global Scholar Award for financial assistance and mentorship.
- Selected among top 35 applicants (out of 1200+) for the Led By Fellowship program 2021 for experiential leadership training.
- Merit based INSPIRE scholarship offered to top 1% students by the Govt. of India.
- Served as the Coordinator of IEEE SIGHT AMU for 2 consecutive years, and the lead of SIGHT Data Science research group.

Umair Ayub

New York, NY | LinkedIn | +1 646-267-6716 | ua2057@nyu.edu

EDUCATION

New York University — **New York, NY**
Master's in Data Science — *Expected Graduation: May 2024*
Relevant Coursework - Data Science, Probability & Statistics, Optimization and Computational Linear Algebra

National Institute of Technology Srinagar — **Kashmir, India**
Bachelor's in Computer Science & Engineering ***(CGPA: 8.61)*** — *Graduated: July 2022*
Relevant Coursework - OOP, Database Management, Java, AI, Deep Learning, Unix & Shell Programming

SKILLS

Programming Languages: Python, Javascript, SQL, C++, Solidity.

Libraries, Frameworks & Tools: Numpy, Pandas, Matplotlib, Seaborn, Tableau, Scikit-Learn, Keras, TensorFlow, PySpark, PyTorch, React, Node.js, Express, AWS, Excel, Microsoft Office, PowerPoint, GitHub.

WORK EXPERIENCE

New York University — **New York, NY**
Teaching Assistant (Programming tools for the Data Scientist Course) — *September 2022 - Present*

- Guiding an undergraduate class of **43** students by explaining the course material and clearing their doubts.
- Assisting the professor in creating & grading assignments, quizzes, and exams.

NIT Srinagar — **Srinagar**
Teaching Assistant (Deep Learning Course) — *August - December 2021*

- Supervised and managed a class of **64** students by providing instructions, leading discussions, and grading exams - attained a class average of **95%** for completed assignments and exams.
- Implemented weekly discussions and mini-projects that encouraged out-of-the-box critical thinking, resulting in enhanced course outcomes.

Feynn Labs — **Remote**
Machine Learning Intern — *August - October 2021*

- Supervised a team of **4** interns to perform market segmentation analysis on the tourism industry of India to find regions that increased the profits by **8%** for a hostel chain.
- Performed Exploratory Data Analysis using **Seaborn** and **Matplotlib** to explain the findings to the senior executives efficiently.
- Designed a Machine Learning model that predicts the profitability of different regions with **80%** precision.

BITS Pilani — **Rajasthan**
Research Intern — *January - July 2021*

- Co-Authored the national bestselling book **"**Machine Learning - A Comprehensive approach**"** which has sold over **4000** copies.
- Designed a Convolutional Neural Network using Python's **TensorFlow** and **Keras** libraries that was aimed at training miniaturized car models for automated driving.
- Collected **50,000+** data samples using **Scrapy** and **Beautiful Soup** for the training and deployment of the Convolutional Neural Network.
- Investigated different e-Healthcare methods, presenting findings in a book chapter available on Taylor and Routledge titled A Taxonomy of e-Healthcare Techniques and solutions: challenges and future directions.

PROJECT

A Novel Deep Learning Approach for Classifying Traffic Signs

- Developed a Deep Learning model using the **(GTSRB)** dataset that used Convolutional Neural Network along with **AutoAlbument** & **Spatial Transformer (STN)**.
- AutoAlbument uses reinforcement learning to automate the search for optimal image transformation policies and STNs results in models which learn invariance.
- The model had **92,397** parameters, was trained in under **3 hours,** and achieved a validation accuracy of **99.86%** which was comparable to the SOTA.

Vandita Goyal

+1551-342-0552 vg2390@nyu.edu Jersey City, NJ 07306

Education

New York University, Graduate School of Arts and Science, Center Data Science — New York, NY, USA
Master of Science in Data Science — May 2024 (expected)

Malaviya National Institute of Technology, Jaipur (NIT Jaipur), — Jaipur, India
Bachelors of Technology (Computer Science) — May 2020

Technical Skills

Languages: Python, SQL, C/C++, HTML/CSS
Core Skills: Data Manipulation and Transformation, Visualisation, Machine Learning, NLP, OOP
Databases: Oracle PL/SQL, MySQL
Python Packages: Pandas, Sci-kit learn, Seaborn,Matplotlib, NLTK, NumPy, SpaCY, Flask

---

Relevant Experience

Data Engineer, Software Development Team — Bengaluru, India
Fidelity Investments — Aug 2020 – June 2022

- Upgraded the back-end application with additional features, enhancing the legacy platform for business users by leveraging Oracle PL/SQL, Informatica, autosys and Python
- Created end-to-end projects by collaborating with multiple teams across the organization involving data collection, clean-up, pre-processing, modeling, testing, and validation as well as experimented with core models, transformers, visualization, time series, and feature engineering. Details mentioned in the Relevant Projects section
- Received “Assist and Automate Enabler” award by Asset Management (India)

Intern, Software Development Team — Bengaluru, India
Fidelity Investments — May 2019 – Jul 2019

- Developed a keyword extraction tool using Word2Vec and Elastic Search
- Used AngularJs and Flask to develop a tool to annotate data (semi-automatic)

Intern, Software Development Team — Noida, India
Nucleus Software — May 2018 – Jun 2018

- Developed a chat bot for NSBT using Rasa NLU and Rasa Core framework written in Python
- Used HTML and CSS to develop front-end, Flask to connect the front-end to the bot and Web Speech API to convert speech-to-text and text-to-speech

---

Relevant Projects

Classification of Fidelity funds into ACWI ex US vs EAFE — Jan 2022 - Jun 2022

- **Associated with Fidelity Investments:** To classify Foreign Equity funds used in Fidelity (SAI) into ACWI ex US & EAFE using machine learning and therefore reduce the overall exposure of each model to emerging market.

Self-Driving ETF — Jul 2021 - Mar 2022

- **Associated with Fidelity Investments:** To create a dynamic fund allocator project that distributes funds across multiple sectors and asset classes depending on market sentiment and economic data.

Prediction of Offer Spread for Corporate Bonds Feb 2021 - May 2021

- **Associated with Fidelity Investments:** To help rank the daily stream of Indication of Interest from our customers, for different corporate bonds.

Depression Analysis Nov 2019 – Jun 2020

- **Associated with NIT Jaipur**: To examine the feasibility of and aim to use different behavioral indicators for depression, consisting of, but not limited to, visual and audio features to design an effective testing model which can be made more accessible than traditional testing methods

---

Additional Accomplishments

- Passed FIT-2 in German from Maxmueller Bhavan
- Organized multiple events as part of Entrepreneurship Development Cell in Blitzschlag ’17, Blitzschlag ’18, Blitzschlag ‘19 and Blitzschlag ‘20 (Graphic Design, Website development, Logistics)
- Volunteered for Cultural Fest Team as a Website developer

ERQIAN (ELSIE) WANG

Jersey City, NJ 07302 | wangeq.elsie@gmail.com | (217) 305-2199 | www.linkedin.com/in/erqian-wang/

EDUCATION

New York University ***Sep. 2022–May 2024***

Master of Science in Data Science

University of Illinois at Urbana-Champaign (UIUC) ***Aug. 2018–May 2022***

Bachelor of Science in Applied Mathematics and Statistics (Double Major) | GPA: 3.99/4.0 (Summa Cum Laude)
Relevant Courses: Statistical Learning, Data Machines and Python, Statistical Modeling, Methods of Applied Statistics, Stochastic Processes, Statistics and Probability, Abstract Linear Algebra

SKILLS

Programming and Tools: Python, R, MySQL, Tableau
Statistics: Linear/Logistic Regression, Ridge/Lasso Regression, Hypothesis Testing, ANOVA
Machine Learning: Decision Trees, Random Forest, XGBoost, k-means, k-nearest neighbors

WORK EXPERIENCE

Data Scientist Intern at PricewaterhouseCoopers (PwC) ***May 2021–Aug. 2021***

Customer Churn Prediction | *Python (pandas, NumPy, matplotlib, seaborn, scikit-learn), SQL*

- Worked cross-functionally with data scientists and product analysts to decrease user churn rate for an e-commerce client using Python.
- Performed data processing (multicollinearity removal, log transformations, standardization, and one-hot encoding) for churn prediction.
- Built and compared Logistic Regression and tree-based models (Decision Trees and Random Forest); tuned hyperparameters for Random Forest; achieved 95% in AUC, 71% in recall, and 94% in precision.
- Evaluated feature importances and assisted in developing a dashboard using Tableau with data automation pipeline ETL for the client to monitor key success metrics on a daily basis, such as DAU, repurchase rate, and customer lifetime value.
- Delivered modeling insights and strategic proposals on churn prevention and promotion to the client.

Data Scientist Intern at Kunlun Health Insurance Company ***May 2019–Aug. 2019***

Health Insurance Fraud Detection | *Python (pandas, NumPy, imbalanced-learn, scikit-learn)*

- Preprocessed data by handling missing values, one-hot encoding, and standardization using Python.
- Utilized SMOTE method to mitigate imbalance in the data by synthesizing new samples for the minority class.
- Constructed XGBoost to detect insurance fraud; increased the recall to 78% (baseline 52%) through oversampling and adjusting sample weights.
- Augmented fraud prevention procedures and projected to reduce losses due to fraud by $3 million annually.

SELECTED PROJECTS

Customer Segmentation with RFM and Clustering | *Python (pandas, NumPy, seaborn, scikit-learn)*

- Aggregated ~500k transaction records into 4k rows for each customer based on the RFM framework.
- Prepared data by standardizing features and removing missing values, duplicates, and outliers.
- Constructed k-means clustering and chose the optimal k value using the elbow method, grouping customers into 4 clusters by their transaction patterns.
- Interpreted clustering results and prioritized customer segmentations for future marketing use.

Email Marketing Effectiveness with A/B Testing | *Python (pandas, NumPy, Plotly, statsmodels)*

- Merged 4 tabular datasets (~1GB) by identifying entity relationships between email campaign datasets deployed to 480k users.
- Defined and computed metrics like email open rate, account linking rate, funding rate, and friction.
- Conducted global hypothesis testing and multiple testing across 24 groups of customers based on their engagement levels; applied Bonferroni correction to alleviate the multiple comparisons problem.
- Visualized conversion funnels to demonstrate 4 fundamental steps in the user journey that lead to funding.
- Made email campaign suggestions based on experimental results.

Xiao, Lu Address: 110 River Dr S, NJ 07310 | Email: xl2473@nyu.edu | Phone:(551)-328-4576

Education

- **NEW YORK UNIVERSITY**, Master of Data Science **(2022.09-2024.05)**
- **NEW YORK UNIVERSITY**, Bachelor of Art in Economics Cumulative GPA : **3.74 / 4.0** **(2018.09-2022.01)**
- Courses: Data Science / Linear Algebra / Statistics / Econometrics / Time Series Analysis / Database / Data Structure

Work Experience

Data Story — Data Analyst Intern **(2021.06-2021.08)**

- Contributed to a whole 2-month online buzz data analysis report set for Budweiser's summer product publicity campaign, including the draft of the scope of work, data collection, data analysis, data reports, and ad-hoc data requirements and created a total revenue of over 50,000 rmb.
- Delivered monthly data reports for more than 5 brands like Innisfree and Starbucks, providing consecutive data trend changes, statistically significant data abnormalities, and corresponding charts and explanations and created a revenue of over 20,000 rmb.
- Implementation of Excel, Python, SQL, Database, Tableau, PowerPoint, and Machine Learning(sentiment analysis).

PricewaterhouseCoopers/PwC ESG — Operation Intern **(2020.12-2021.02)**

- Designed and conducted a set of A/B tests about teaching and curriculum quality for more than 120 students at 3 different workspaces, finally optimized 3 teachers and 1 course structure.
- Organized 3 large ceremonies by designing time flow charts and allocating human resources and increased working efficiency by 30%.

Projects Experience

Public class projects of Deep Learning, Hung-yi Lee **(2022.01-2022.06)**

- Performed a frame-wise phoneme classification model using pre-extracted MFCC features, achieving 78% accuracy.
- Performed an image classification model of 11 categories of food using CNN and improved the model's performance by implementing cross-validation and data augmentations such as mix-up and transforms, achieving 82% accuracy.
- Performed a multi-speaker classification model from VoxCeleb2 dataset by implementing transformer and attention, achieving 71% accuracy.
- Developed an anime face generation model by implementing GAN.

JoinQuant & Soochow Securities **(2022.01-2022.03)**

- Developed and back-tested a stock transaction model, built up a portfolio of better securities using a multi-factor rating strategy, and implemented it onto real a transaction platform, outperforming the stock market by a 17.3% annual return rate through 5 months.

New York University, Pascal Wallisch **(2021.01-2021.06)**

- Analyzed 8 middle schools' admission rates in New York out of over 20 factors such as Standardized Test score, distributions of different students, and facilities of schools.
- Implemented Principal Component Analysis, correlation analysis, significance tests, and regression to summarize key factors that affect admission rate and provide government 4 empirical suggestions for improvement.

Skills & Interests

- Python (Numpy / Pandas / Matplotlib) / MySQL / Java / R / C# / C / Excel / Tableau / PPT / Statistics
- Regression Model (Linear / Logistic / Multi-variable)
- Machine Learning (KNN / K-means / SVM / Random Forest / Bagging & Boosting / PCA)
- NLP (Transformer / Attention) / Deep Learning (Pytorch / CNN / RNN / GAN)
- Economic Models / Time Series Analysis / Quantitative Security Strategies & Models

XINYUE MA

xm618@nyu.edu| +1(617)-860-9142 | Brooklyn, NY | www.linkedin.com/in/xinyue-ma1126/ | www.github.com/xinyuem7

EDUCATION

New York University, Graduate School of Arts & Science — New York
Master of Data Science — Sep 2022 - May 2024
Relevant Courses: Optimization and Computational Linear Algebra, Machine Learning

Boston University, College of Arts and Sciences — Boston, MA
Bachelor of Science, Mathematics (Statistics) | ***Minor**: Comp. Science* | ***GPA:** 3.84/4.00* | *Magna Cum Laude, Dean's List* — May 2022
Relevant Courses: Linear Models, Stochastic Processes, Statistics, Probability, Time Series & Forecasting, Data Science in R

RESEARCH EXPERIENCE

Experimental Approaches for Tuning NIC Registers, Boston University — Boston, MA
Undergraduate Assistant — Sep 2020 - Dec 2021

- Conducted exploratory data analysis (EDA), feature selection and feature engineering on a dataset with 120K+ records from 2 NIC Registers using Pandas & NumPy in Python, identified & treated data anomalies, explored & visualized the pairwise relationships in the data using Matplotlib.
- Built a regression model using Sci-kit learn in Python to predict the energy-delay product; implemented gradient descent to optimize the model parameters, which resulted in an adjusted R-squared value of 0.75.
- Utilized the regression model to predict the energy-delay product with sets of new inputs, which helped to avoid running time-consuming and expensive experiments.

PROJECTS AND LEADERSHIP

Analysis of the US Saving Rate from 1960 to 2020 | *R, Regression* — Jan 2021 – May 2021

- Cleansed & processed past 60-year financial spend & saving records & 10+ macro-economic factors in US using dplyr in R, explored the data with trendlines, pie charts & bar graphs using ggplot2.
- Developed a linear regression model with glm in R to determine the impact of macroeconomic factors on American saving & spending behavior, summarized the project insights in webpage. (https://economatrix.netlify.app/)
- Detected multicollinearity with a correlation plot & implemented Mallow's CP Value in R (leaps) to identify the optimal linear regression model, achieved an adjusted R-squared value of 0.81.
- Designed & created an interactive webpage using Shiny which allowed users to explore & customize various linear regression models & evaluate efficacy of the custom model using 3 statistical metrics. (https://economatrix.shinyapps.io/economatrix/)

Mobile Application 'Timato' Development | *Dart, SQL, Agile, Product Development* — May 2020 – Aug 2020

- Developed a mobile application 'Timato' that implemented the Pomodoro Technique for users to them improve daily work/study efficiency, published the app on the App store & Google Play resulting in 100+ downloads.
- Led the conceptualization of the functionalities & designed the UI using Figma, built the frontend using Dart & SQLite as backend database with an Agile approach, produced custom SQL queries to realize each of the app functionalities.
- Conducted user acceptance testing for the app to ensure usability & eliminate bugs prior to launch.

WORK EXPERIENCE

Program Data & Evaluation Intern **- 826 Boston**, Boston, MA — Jan 2022 - Jul 2022

- Created & deployed an end-of-year survey to evaluate students' learning progress, gathered & cleansed 300+ student responses.
- Analyzed survey response data and visualized trends & patterns using pivot table & pivot chart in Excel to quantify the impact of the educational program & justify continued grant funding.
- Prepared & published weekly newsletter to 50+ internal stakeholders with recent updates in program highlights, professional development opportunities, and student resources.

Data Analyst Intern **- University Service Center,** Boston University, Boston, MA — Sep 2021 - Dec 2021

- Consolidated, cleansed & transformed data for students who withdrew or took leave from university in past 5 years with SQL.
- Performed quantitative & qualitative analyses using Python Pandas & NumPy, imputed missing data & treated outliers, visualized trends & patterns in the data with Matplotlib.
- Identified & showcased 3 key actionable insights & opportunity areas for the university to increase student retention.
- Presented data-driven recommendations to improve student retention to department heads, leading to increase of department-wide funding by 15% and launches of 2 programs dedicated to mental health and community-building activities for students.

Consultant Assistant Intern **– Capgemini,** Beijing, China — Jul 2021 - Aug 2021

- Designed Entity-Relationship diagrams to showcase database-design capabilities to 5+ prospective clients.
- Analyzed visitor behavior in 15+ amusement parks & ticket sales performance post new ticket system implementation using Tableau, prepared quarterly reports to showcase the impact of the new system on sales & presented to client leadership team.

Teaching Assistant for Intro to CS II **- Computer Science Department**, Boston University, Boston, MA — Jan 2020 – May 2022

- Conducted weekly lab sessions, resolved students' queries during weekly office hours & graded exams for 120+ students.
- Guided & mentored students on difficult CS concepts with additional study material.

SKILLS

Programming Languages: Python, SQL, Java, R
Analytical Tools: Tableau, Dart, Microsoft Excel, PowerPoint
Techniques: Machine Learning, Data Mining & Analysis
Languages: English (Advanced), Mandarin (Native)

Yajie (Zoe) Xiao

yx1750@nyu.edu | (551)349-3250
GitHub | LinkedIn | Personal Website

EDUCATION

New York University — New York City, United States
M.S. in *Data Science* — Sep 2022 - May 2024

New York University Shanghai — Shanghai, China
B.S. in *Data Science*, Minor in *Interactive Media Arts* (Dean's List) — Sep 2018 - May 2022
Study Away at New York University Abu Dhabi from Sep 2021 to May 2022

SKILLS

Languages: Native in Mandarin Chinese; Proficient in English
Data Science: Python and Machine Learning Toolkit (NumPy, Pandas, SciPy, Matplotlib, Scikit-learn), MySQL
Software Engineering: HTML, CSS, JavaScript (React, React Native, D3), Python Flask, Git
Software Skills: Jupyter Notebooks, Spyder, LaTeX, PowerBI, Adobe XD, Figma, QGIS, Stata, SPSS

WORK EXPERIENCE

Fabernovel (Strategy) — Shanghai, China
Data Analyst Intern — May 2021 – Aug 2021

- Built and maintained 5 key dashboards to track Consumer Relationship Management metrics including Customer Lifetime Value, Customer Churn Rate using PowerBI
- Designed Chinese Natural Language Processing tools in Python and implemented sentiment analysis, topic extraction, and word frequency statistics to analyze consumer feedback, quantified the overall customer experience for 5 projects
- Drafted 5 major data insight reports, summarizing key metrics and presenting visualizations for 3 client companies

FNZ (FinTech) — Shanghai, China
Engineer Intern — Jun 2021 – Aug 2021

- Assisted project deliverables and UX testing in an agile team, reporting to 3 Asset Management Firms
- Built 20+ UI for finance app across iOS and Android, including modules such as Investments, Documents, Transactions
- Designed interactive analytics dashboard with UI design team, led a team of 5 to build the dashboard

PROJECT EXPERIENCE

Using Satellite Image Data to Understand Market Access — New York University Abu Dhabi
Visiting Student Research Assistantship Award — Sep 2021 – May 2022
The project exploited a large volume of satellite image data in topography to provide a new measure of market access for evaluating the contribution of an area's interconnections to its level of economic activity. The derived measures were based on theories in linkage analysis associated with the Leontief inverse matrix and were created using Python and QGIS.

User Behavior Analysis on Social Network Under Data Privacy Risks — New York University Shanghai
Summer Dean's Undergraduate Research Funding (DURF) Scholarship — Jun 2021 - Aug 2021

- Developed a Python web-scraping program to extract data of 10,000+ rows from Weibo (Chinese social platform)
- Ran linear regression models between user satisfaction index and related factors including user engagement, post count, and posting time, generated correlation analysis and model summary statistics

End-to-End Chinese Carplate Detection and Recognition System — Mar 2021 - May 2021
The Computer Vision project utilizes Python to detect and recognize Chinese carplate from 1000+ CCTV images using deep neural networks LeNet-5 and YOLOv5, resulting a Chinese character recognition model with 90%+ accuracy. GitHub

Online Flight Ticket Reservation System — Apr 2021 - May 2021

- Identified key entities in online booking to build a flight ticket reservation system, allowing customers, booking agents, and airline staff to purchase tickets, view flight status, and view commission for agents; designed E-R diagram
- Implemented front-end application using HTML, CSS, JavaScript, and utilized MySQL to connect back-end Python Flask App and relational database . GitHub

WANYI YANG

yangwanyi77@gmail.com | (1) 646 575 6976 | www.linkedin.com/in/wanyi-yang1

EDUCATION

New York University New York, US Sep. 2022 – May 2024
Master of Science in Data Science
Relevant coursework: Machine Learning; Databases; Linear Algebra; Probability; Data Structure.

New York University Shanghai Shanghai, China Sep. 2018 – May 2022
Bachelor of Science in Business and Finance & Data Science
Study away at NYU Stern School of Business (Sep. 2021 - May 2022)

SKILLS

Computer skills: Python, R, SQL, JavaScript, HTML/CSS, Tableau, Capital IQ, Bloomberg, Excel, PowerPoint(think-cell)

PROJECTS

Customer Retention at HSBC | Python Jan 2022 - May 2022

- Integrated 116K customer structured data from multiple internal and external sources (13 tables).
- Explored historical data to create visualization dashboard (Tableau, and ggplot in Python).
- Identified customer retention by using 4 classification models including logistic regression, Support Vector Machine, XGBoost and Neural Network, reached accuracy over 95%, retained 15% inactive user to active user by simulative test.
- Collaborated with cross-functional teams including customer service, marketing team to influence future product roadmap, delivered retention strategies of inactive customers to the leadership.

COVID-19 Cases Prediction – A Time Series Prediction Using Machine Learning Approach | Python Dec 2020

- Build a 3000 covid-19 cases dataset from 7 countries, with their country features.
- Implemented basic models like ridge regression, regression tree, KNN & random forest to predict future covid-19 cases. Applied single and multiple LSTM models to the time series data, tuned the hyper-parameters.
- KNN performed best with an MAE of 0.0134 to predict the number of cases in upcoming days.

Online Air Ticket Reservation System | SQL, Python, JavaScript, HTML, CSS Mar 2020 – May 2020

- Developed a system that allows customers to search and purchase tickets and airline staff to manage flights.
- Designed the UI wireframes & developed a front-end application using JavaScript and HTML and connected its functionalities to the back-end database using Python MySQL package.

EXPERIENCE

***Post Investment Analysis Intern,* Tencent Games,** Shenzhen, China Oct. 2021 – Feb. 2022

- Developed IC memos for 2 targeted video game company in Poland with holistic analysis of company's potential.
- Built 6 regional video game benchmarks & comps using Capital IQ to monitor 27 portfolio companies' market performance.
- Diagnosed the global gaming industry with investment methodology, analyzing top 50 video game company's financial reports, game pipelines, company structure, turnover rate, M&A transactions, etc.
- Evaluated 27 companies in Tencent's investment portfolio by IP evaluation, game pipeline projection, financial analysis.

***Consulting Intern,* AlixPartners,** Shanghai, China Apr. 2021 – Aug. 2021

- Modeled saving analysis using Excel to do cost reduction analysis for a UK retail platform, saved over $1.3 billions after optimization.
- Processed various sample request orders and managed suppliers to ensure work efficiency.
- Conducted desktop research on autonomous truck, electronic accessory, fresh supermarket, prepare reports for new bids.
- Screened global suppliers, make cold calls, and negotiate import/export (port, tax, pay term) details with them.

***Marketing Strategy Intern,* SenseTime Technology,** Shanghai, China June 2020 – Oct. 2020

- Proposed and organized SenseTime AI summer workshop, communicated for cross-departmental cooperation.
- Managed over hundreds of orders using CRM system and assisted in semi-annual sales performance calculation.
- Arranged multiple projects (eg. annual security meeting).

Yuhao (Bill) Liu

(617)710-8647 | New Jersey | yl5308@nyu.edu | https://github.com/LYH36 | linkedin.com/in/yuhao-bill-liu-088519177

EDUCATION

New York University **New York, NY**

Master of Science in Data Science Sep 2022 – May 2024

- **Coursework:** Optimization and Computational Linear Algebra, Programming for Data Science

Boston College **Chestnut Hill, MA**

Undergraduate Degree, B.A. in Economics, B.A. in Mathematics | ***GPA:*** *3.8/4.0* Aug 2018 – May 2022

- **Honors:** Dean's List First Honors
- **Coursework:** Statistics, Probability Theory, Linear Algebra, Algorithm, Machine Learning, Data Science, Optimization

SKILLS

Tools: Python, Microsoft Excel, STATA, SPSS, SQL, Java, Tableau, JMP, R, BigQuery, HTML, MongoDB, JavaScript
Techniques: Data Mining, Machine Learning, Statistics, Data Analytics, Data Visualization, Mathematical Modeling.

PROFESSIONAL EXPERIENCE

P&G **Beijing, China**

Data Science Analyst – Intern Jul 2021 – Aug 2021

- Gathered **~20k** records of diaper packaging data, captured **15** key business metrics based on consumer tests in Japan, cleansed & transformed the data in **Excel**, summarized the insights for decision-making by baby care department.
- Applied linear regression in **JMP** to quantify the impact of compactness on length of package with **R-squared of 0.93**.
- Envisioned & created the **1st** globally used business model for diaper packages, maximized diaper's softness and quantity, improved consumer perception scores by **20%** and adopted as the standard for global package design.

MaiMai **Beijing, China**

Data Analyst – Intern (User Growth Team) Feb 2021 – Apr 2021

- Cleansed & processed **100k+** user transactions using **Google BigQuery** and conducted exploratory **Big Data** analysis, implemented user's retention rate regression model using **Scikit-learn in Python**, achieved an **RMSE of 0.067**.
- Designed daily refreshed **Power BI** Dashboard with 8 key metrics & integrated automatic alerts for real-time decision-making.
- Analyzed spurts in user growth to identify root causes, developed user personas applied on user activity & demographic features and provided data-driven recommendations for marketing & product decisions.

Gerson Lehrman Group **Beijing, China**

Financial and Data Analyst – Intern Sep 2020 – Jan 2021

- Extracted relevant information from **10M+** historical transactions using **SQL** scripts, analyzed the data with summary statistics & pivot charts in **Excel**, identified anomalies & patterns in the data & presented insights to manager.
- Developed & monitored data quality in the database with **Python Pandas**, visualized data quality report with **Seaborn** plots & achieved **98%** data quality compliance.
- Defined **12+** employee performance metrics and computation logic, transitioned to a data-driven appraisal system & reduced Chinese section's labor costs by **10%**.

KPMG **Online**

Business Analytics Consulting Virtual Internship Jul 2020 – Jul 2020

- Identified **15+** key business metrics for Sprocket Central, presented quantitative analysis in **Tableau** dashboard.
- Assessed data quality for dataset of **20,000** customers with **30+** fields, generated summary statistics, imputed missing data and treated outliers, conducted data deduplication & normalization, improved overall data quality by **24%**.
- Applied **k-means clustering** for customer segmentation & targeted high-value customers based on 8 customer demographic & transaction attributes, optimized current transactional data model.

PROJECTS

Boston College COVID-19 Prediction **Chestnut Hill, MA**

Project leader Apr 2020 – May 2020

- Conducted COVID-19 prediction using **Python Pandas**, visualized prediction model with **Matplotlib plots**.
- Applied **TensorFlow** and **Scikit-learn in Python** to predict the peak death rate date with **R-squared of 0.87.**

JINYANG LIU

Jersey City, NJ | jl14310@nyu.edu | (551) 344 -1565

EDUCATION

New York University, New York, NY — Expected May 2024
Master of Science in Data Science
Current Coursework: NLP with Representation Learning, Machine Learning in Finance

College of William and Mary, Williamsburg, VA — September 2018 – December 2021
Bachelor of Science in Standard Mathematics, Minor in Computer Science
Cumulative GPA: 3.84/4.00; Major GPA: 3.90/4.00
Coursework: Databases, Data Visualization, Machine Learning, Linear Algebra, Statistics, Probability, Statistical Data Analysis, Numerical Analysis, Operations Research Stochastic Models

TECHNICAL SKILLS

Coding Languages: Python, SQL, R, C++, C, HTML, CSS
Software and Tools: Tableau, MS Office, MATLAB, Latex

RELEVANT EXPERIENCE

DisinfoLab | Williamsburg, VA
Technical Analyst — June 2021 – August 2021

- Developed an automated data collection code using Python and Twitter API to collect and clean inauthentic Twitter accounts information and their activity
- Fetched and fed 20,000 data points in 2 weeks to the team's machine learning model
- Performed sentiment analysis on tweet texts with NLTK, the natural language processing library

AidData Transparent Development Footprint Team | Williamsburg, VA
Senior Research Assistant — October 2019 – May 2021

- Researched and collected data on underreported Chinese overseas development finance programs. Wrote analysis reports on the financing and the implementation of each project
- Compiled and quality assured the data of 985 projects into the team's database, covering 23% of the projects towards Southeast Asia
- Collaborated with a team of 25 with weekly meetings and reports

ACADEMIC PROJECTS

Interactive Maps of Famous Artworks (python, HTML, Tableau) — April 2021
Web scraped and cleaned the data of 300 artworks; Designed a website with interactive maps to show their details and locations in popups

Broadway Earnings Data Analysis (R, Excel) — November 2020
Analyzed and built a stepwise regression model on the data of 31,000 weeks of Broadway shows. Presented to a class of 40 and completed a paper

Relational Database Project (SQLite) — November 2021
Created a relational database to organize 50,000 sample retail data records