

Tushar Gupta

tg2749@columbia.edu | [linkedin.com/in/tushargupta95](https://www.linkedin.com/in/tushargupta95) | (+91)-9547010744

EDUCATION

Columbia University

M.S. in Computer Science, Machine Learning Track

Courses: N.L.P, Intro to Databases, Cloud computing & Big Data, Practical Deep Learning Systems

New York, US

May 2022

Indian Institute of Technology, Kharagpur

Dual Degree (B.Tech. & M.Tech.)

Minor in Computer Science and Engineering

Kharagpur, IN

May 2018

PROFESSIONAL EXPERIENCE

Bayer U.S. LLC | Consumer Healthcare

New Jersey, US

Senior Data Scientist | Analytics & Forecasting

June'23 – Dec'23

- Spearheaded a **52-week** retail sales forecasting system for the U.S. product portfolio worth **~3B \$** across 15 brands & 7 retailers for the Demand Planning team
- Engineered robust system to capture market anomalies via features such as weather, promotions, store inventory, competitor sales, illness incidences and more
- GB Trees and LSTM-based time-series models achieved a Lag 3 accuracy between **70%** and **85%** across 200+ SKUs with a maximum **5pp** seasonal deviation
- Implemented automated model retraining triggers and weekly prediction generation in Databricks, resulting in efficiency improvements worth 10 hours/week
- Introduced a business strategy simulator with the Category, Brands and Supply, enabling teams to build interpretable scenarios using standardised human inputs.

Data Scientist | Analytics & Forecasting

June'22 – May'23

New Product Sales Estimation:

- Launched predictive models to estimate new sales opportunities (1-year sales) with **93%** accuracy for upcoming Bayer Over-the-counter (OTC) launches.
- Utilised regression-based models on POS, Distribution and Brand penetration data of similar product introductions in the market within the past 5 years
- Partnered with 14 Innovation team managers on 38 product forecasts & market positioning analysis, translating to **1.7 M \$** in savings for the team
- Developed quarterly Long-term (5 - Year) OTC forecasting system across 50+ categories, used by sales teams in building stronger relationships with Walmart
- Managed an intern over Summer'23 who improved A&F team's flu-season forecasting accuracy by including southern hemisphere trends in modelling pipeline
- Consistently recognised for talent receiving Top Performance & R4B awards in 2023 for leading innovation within the team

Gartner Pvt. Ltd.

Gurgaon, IN

Senior Business Associate | Product Development

Mar'20 - Nov'20

- Managed product design and development of **Python & SQL** data pipelines for an org-wide live client prioritisation tool used by **300+** service associates
- Streamlined bug resolution processes with issue support teams and designed data monitoring tools in Python to reduce monthly reported issues by **38%**
- Collaborated with multiple scrum teams and product stakeholders to build business roadmap and standardised client outreach metrics across 7 verticals

Quantitative Analyst | Data Science

June'18 - Feb'20

Early-Risk Prediction System:

- Identified early-client risk prediction indicators via feature analysis on client usage indicators to build statistically guided success metrics for service associates
- Recorded **90%+** recall in identifying churned customers via RandomForest models based on risk prediction features designed at multiple touchpoints in contract Content recommendation for Unengaged users:

- Innovated a word-vector-based content matching algorithm for suggesting research documents to low-engaged users based on engagement of organisation peers
- Achieved **80%** hits in *top-10* results for various user profiles created across markets and industries. Piloted **Flask** service with an outreach team that integrated risk predictions with content suggestions to gather associate feedback

Client Feedback Classification:

- Launched a text classification service using Multinomial Naive Bayes to publish root causes of service dissatisfaction from client surveys. Achieved **F1-score** of **0.75** and transformed model results into key actionable suggestions. Implemented a PowerBI and Python pipeline to reduce monthly **reporting time** by **50%**
- Utilised **Sklearn** for model development, **stats model** Python API for experimentation, **SQL** for data extraction and **AWS** cloud VMs for computation

Edge Networks

Bangalore, IN

NLP Research Intern

May'17 – July'17

- Innovated a Bi-LSTM & ConvNet strategy in Pytorch for domain-independent **resume classification**. Applied Attention over embeddings and increased top-10 precision by **30%** points. Created on-demand APIs using **Flask** for the above models to facilitate user validation and product integration

ACADEMIC PROJECTS

Stock Price prediction via Tweets, Big Data Analytics (CU)

Sep '21 – Dec '21

- Tabulated performance of LSTM and statistical regression models for predicting the next 3-day stock price trend using N-day moving averages of historical prices and twitter sentiments in top 5 tech companies. Utilized **BigQuery** for data storage, **Pyspark** for data ETL & **Airflow** for automated pipelines.

Smart Image Gallery, Cloud Computing and Big Data (CU)

Jan '21 – April '21

- Collaborated with a team of 3 people to design and execute a photo gallery application with image upload and **custom search functionality** using AWS Elasticsearch, Lambda and Rekognition services. Employed AWS CodePipeline for automated change deployments (**CI/CD**)

Indian Institute of Technology, Kharagpur

Aug '17 – May '18

- Awarded top Master's thesis at the Annual Convocation, 2018, for work on stochastic optimisation and model predictive control in the crystallisation domain

TECHNICAL SKILLS

- Programming:** Python, SQL, C, R, HTML5/CSS, GoLang
- Frameworks:** Pytorch, Keras, PySpark, NLTK, NumPy, Nvidia DeepStream, Pandas, Flask, scikit-learn
- Software / Platform:** DataRobot, Databricks, Snowflake, AWS, Google Cloud (Dataproc, BigQuery, Apache Airflow), Linux, Git, Power BI, MATLAB