# Tushar Gupta

(91) 95470-10744 tushariitkgp14@gmail.com linkedin.com/tushargupta95 github.com/tushargupta14

## EDUCATION

**Columbia University** May 2022
*M.S. in Computer Science, Machine Learning Track* *New York, US*

**Indian Institute of Technology, Kharagpur** June 2018
*Dual Degree (B.Tech & M.Tech) with Minor in C.S.E* *Kharagpur, IN*

## EXPERIENCE

**Bayer U.S Consumer Health** Whippany, US
*Senior Data Scientist* *Jun 2023 – Dec 2023*

- New Product Sales Forecasts: Built regression-based models to predict incremental sales opportunity for upcoming Bayer Over-the-counter (OTC) launches, with features based on past similar product introductions in the market.
- Generated **1.7 M$** in savings with an in-house forecasting engine from 38 product forecasts in partnership with 14 innovation managers. Models tested **93%** accurate over a 6-month period.

*Data Scientist* *Jun 2022 – May 2023*

- **SKU Demand Forecasting System**: Developed a Gradient Boosted trees & LSTM based time-series forecasting pipeline for Bayer OTC drug portfolio worth $\sim 3B\$$ (15 brands) to predict 1-year future demand at weekly granularity
- Ensured broad market coverage via features such as weather, promotions, store inventory, competitor sales and illnesses to achieve a Lag 3 accuracy ranging between **70%** – **85%** across 200+ SKUs with $\pm$**5pp** seasonal deviations. Integrated automated accuracy based re-training triggers on Databricks to significantly diminish weekly model maintenance effort
- Consistently recognised for talent receiving Top Performance award in 2023 for leading innovation within the team

**Gartner Research & Advisory** Gurugram, IN
*Senior Business Associate* *Mar 2020 – Nov 2020*

- Led product development of Python & SQL data pipelines for an org-wide live client prioritisation tool used by 300+ service associates daily. Reduced monthly reported issues by **38%** via Python based automated data quality checks

*Quantitative Analyst (NLP)* *Jun 2018 – Feb 2020*

- Engineered **Early-Risk Prediction indicators** on client usage patterns to build statistically guided success metrics for service associates. Recorded **90%+** recall in identifying churned customers via RandomForest models at multiple contract touchpoints
- Innovated a word-vector-based **Textual Content recommendation algorithm** for suggesting research documents to **unengaged users** via collaborative filtering. Achieved **80%** hits in top-10 results for user profiles across markets and industries
- Launched a **Client Feedback Text Classification** service using Multinomial Naive Bayes to publish root causes of service dissatisfaction from client responses with an F1-score of 0.75. Reduced insight analysis time from survey text data by **50%**

**Edge Networks** Bangalore, IN
*NLP Research Intern* *May 2017 – Jul 2017*

- Innovated a Bi-LSTM & ConvNet strategy in Pytorch on resume embeddings for text classification into suitable job-types. Implemented Attention over embeddings to improve top-10 precision by 30pp. Created on-demand APIs using Flask.

## ACADEMIC PROJECTS

**Stock Price Prediction via Tweets** | *LSTM, Pyspark, BigQuery, Python, Airflow, Streamlit*

- 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

**Smart Image Gallery** | *AWS Lambda, CodePipeline, Elasticsearch, Lambda, Rekognition, Python*

- Designed and executed a photo gallery application with image upload and custom search functionality using AWS services

## TECHNICAL SKILLS

| | |
|---|---|
| Programming Languages: | Python, SQL, C, HTML5/CSS, GoLang |
| Frameworks: | Pytorch, Keras, Pyspark, Flask, MySQL, MongoDB, NumPy, Matplotlib, scikit-learn, NLTK, fastText, BERT |
| Tools & Technologies: | Datarobot, Snowflake, Databricks, AWS Sagemaker, Google Cloud (Airflow, BigQuery), Git, Linux, LaTeX, Agile/Scrum, PowerBI, Pandas, Streamlit |