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   Whippany, US
Senior Data Scientist   *Jun'22 – Dec'23*

- **SKU Demand Forecast Engine**: Developed automated time-series forecasts and feature ETL pipeline for a 3B$ OTC product portfolio (200+ SKUs) to enable robust supply planning and accurate sales estimation. Gradient Boosted trees & LSTM architectures achieved avg. **80%** forecast accuracy for 1-year future retail sales with $\pm$**5pp** seasonal deviations.
- Productionized model inference on DataRobot with automated retraining triggers on Databricks & Snowflake to eliminate weekly model maintenance efforts.
- **Product Launch Forecasts**: Utilized regression-based models to predict new sales opportunity size for upcoming Bayer product launches. Generated **1.7 M$** in savings with an in-house sales forecasting tool in collaboration with 14 innovation managers.
- Awarded Top Performance & R4B (Remarkable for Bayer) awards in 2023 for delivering AI solutions and leading innovation.

Gartner   Gurugram, IN
Quantitative Analyst (NLP)   *Jun'18 – Nov'20*

- **Client Churn Prediction:** Engineered risk-indicators based on user engagement patterns to identify likely-to-cancel premium customers. Recorded **90% Recall** via RandomForest models trained at multiple contract touchpoints to facilitate early detection.
- **Research-content recommendation:** Innovated a word-vector-based algorithm for boosting content viewership using a collaborative filtering approach. Achieved **80%** click-through-rate in top-10 suggestions for users across markets and industries.
- **User Feedback Classification:** Launched a Flask service to publish root causes of service dissatisfaction from client responses with an F1-score of 0.8. Integrated automated PowerBI based data visualization pipeline reducing insight analysis time by **50%**.

Edge Networks   Bangalore, IN
NLP Research Intern   *May'17 – Jul'17*

- **Skill Indentification:** Innovated a Bi-LSTM architecture in Pytorch over recruiter job-descriptions to segregate skill-based keywords independant of industry domain. Recorded a macro avg. precision of 96% on 5 skill-categories.
- **Resume Mapping:** Implemented RankNet with attention over resume embeddings for mapping candidate resumes to similar job titles and increased top-10 precision by **30pp**. Created on-demand APIs using Flask to facilitate robust validation.

ACADEMIC PROJECTS

Map-Reduce Server | *Go, Distributed Systems*
- Built a key/value based Map-Reduce service focusing on fault tolerance

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, GoLang
Frameworks:	Pytorch, Keras, Pyspark, Flask, MySQL, MongoDB, scikit-learn, fastText, Pandas
Tools & Technologies:	Datarobot, Snowflake, Databricks, AWS, Google Cloud, LaTeX, PowerBI, Streamlit