(716) 903-7512
Buffalo, NY
sagarwork18@gmail.com

Sagar Jitendra Thacker
Data Scientist / MLE

sagarthacker.com
github.com/sagar118
linkedin.com/in/sagar-thacker

Data Scientist with a **Master's** degree in Computer Science & over three years of experience designing and deploying **Machine Learning**, **NLP**, and **Deep Learning** applications, seeking for Data Scientist or Machine Learning Engineer roles.

TECHNICAL EXPERIENCE

Research Scientist, *University at Buffalo*   Mar 2023 — Present
- Research and develop a classifier to detect racist language in social media text. (ongoing)

Graduate Grader Assistant, *University at Buffalo*   Jan 2022 — May 2022
- Evaluated and provided constructive feedback on student assignments, facilitating problem-solving.
- Conducted regular office hours to resolve students' inquiries related to grades, assignments, & course material.

Data Scientist, *FedEx Express*   Jan 2020 — Apr 2021
- Utilized **NLP - Sentiment Analysis** with **PyTorch** to extract sentiments from customer feedback & ticket data, improved support efficiency by **30%**. Deployed model as **Dockerized web service** on **Flask**, utilizing **RESTful API** calls.
- Optimized lead conversion pipeline utilizing **Logistic Regression** & **XGBoost** ensemble model to identify client proposals with highest likelihood for success.
- Developed **hierarchical classifier** with **Logistic Regression** and **SVM** to direct customer tickets to relevant departments for effective resolution, enabling **65%** reduction in resolution time.
- Constructed automated **ETL pipelines** to extract competitor rates from PDFs, reducing manual efforts by **90%**.
- Designed performance tracking dashboards with **Tableau** & **Power BI**, increased operational efficiency by **15%**.

Assistant System Engineer, *Tata Consultancy Services*   Jul 2018 — Nov 2019
- Developed test scripts with **Java** & **Selenium** to simulate testing process in web apps, reduced manual labor by **75%**.
- Modeled customer/software interaction in **ARD** for automatic test script generation, reducing manual effort by **80%**.

EDUCATION

Master of Science, Computer Science, *University at Buffalo, The State University of New York, 4 GPA*   Feb 2023
Bachelor of Engineering, Computer Engineering, *University of Mumbai, 8.69 GPA*   Jun 2018

PROJECTS

Social Media Mining for Health Monitoring
- Led a team of 2 to utilize **BERT**-based **transformer** models for addressing healthcare challenges in social media texts.
- Developed **Named Entity Recognition (NER)** & **multi-class classification** models in **PyTorch** to extract & classify adverse effects of medicines and normalized text representation leading to **17%** improvement in scores.
- Enhanced automatic MedDRA code normalization with **SciBERT** text embeddings and cosine similarity.

Trip Duration Prediction
- Created an **XGBoost** regression model to predict the duration of NY Taxi trips. Registered model on **S3** using **MLflow**.
- Deployed model (Docker container) on **AWS ECR** with **AWS Kinesis** and **AWS Lambda** as inference pipeline.

NYC-Taxi Data Analysis
- Designed GCP infrastructure using **Terraform** for **Google Cloud Storage** and **Google BigQuery**.
- Developed **dockerized Prefect** pipelines for extracting and loading data from web to cloud storage to BigQuery.
- Deployed, tested & documented **dbt** models to transform BigQuery data on dbt cloud, visualized data on **Data Studio**.

ACHIEVEMENTS

Secured **second place** in the H2O World India Hackathon held in April 2023

SKILLS

Languages: Python, Java, C++, JavaScript, SQL **Databases**: PostgreSQL, MongoDB, MySQL, SQLite
Cloud: AWS (Beanstalk, EC2, S3, Kinesis, ECR, Lambda, IAM), Azure, GCP (Cloud Storage, BigQuery)
Others: Docker, Tableau, Power BI, PySpark, Spark, Linux, Git, Statistics, Bash, Terraform, A/B Testing, Prefect, dbt
ML: scikit-learn, Keras, TensorFlow, PyTorch, SciPy, MLflow, WandB, spaCy, NLTK, XGBoost