Shashank Gupta

+1-862-318-1725 | Kearny, NJ 07032
shashank.and.gupta@gmail.com | linkedin.com/in/shashank47 | github.com/Sha661nk | kaggle.com/wrecked22

Education

Rutgers University, New Brunswick **Sep 2022 – Dec 2023**
Masters in Information Technology & Analytics, (Received Merit Scholarship) *GPA: 4.0/4.0*

Indian Institute of Technology, Kanpur **Jul 2015 – May 2019**
Bachelor of Technology in Electrical Engineering *GPA: 3.4/4.0*

Technical Skills

Languages: Python, R, SQL, Cypher (CQL)
Tools: Jupyter Notebook, R Studio, SIMCA, PostgreSQL, MySQL, Power BI, Oracle, AWS Sagemaker, AWS Lambda, AWS DynamoDB, Tableau, OSIsoft PI, Neo4j, MS Excel, Git & Github
Frameworks & Libraries: Pandas, NumPy, scikit-learn, PySpark, Heroku, Flask, pytorch, Matplotlib, NLTK, SpaCy, BeautifulSoup, ggplot2, CV2, Seaborn, plotly, pytesseract, Gensim, Keras, TensorFlow, OpenCV

Work Experience

Sanofi US **Jun 2023 – Present**
Data Science PAT Co-Op (Skillset: Chemometric, Machine Learning, Data Science, Python, SQL) *Boston, MA, US*

- Implemented Machine Learning and Chemometric techniques to create a generalized, automated pipeline for predicting process quality attributes using the high dimensional Raman Spectroscopy data.
- Developed a spectral data pre-processing pipeline, utilizing Grid-Search, PLS model, and custom metrics to identify and apply optimal set of data pre-processing techniques ultimately reducing RMSECV by 20%.
- Performed Feature Selection by leveraging the Wet-Lab experiments and employing three feature selection techniques, including Filter, Wrapper, and Embedded methods.

I-DSLA - Rutgers University **Apr 2023 – Present**
Data Science Researcher (Skillset: Generative AI, Neural Networks, R, Python) *New Brunswick, NJ, US*

- Evaluated three generative models including Variational Auto Encoders (VAE), CT-GAN, and Bayesian Network to assess their effectiveness in preserving the missing data distribution in generated synthetic data.
- Developed a data evaluation pipeline to assess the quality of synthetic data using t-SNE plots, Wasserstein Distance, KL Divergence, REM, RESD, and PCD metrics.
- Conducted a comprehensive evaluation on Generative models, revealing effectiveness of Bayesian Network over CT-GAN in generating synthetic data and notably preserving the missing data distribution by a significant margin.

LTI-Mindtree **Jul 2019 – Sep 2022**
Senior Data Scientist (Skillset: Machine Learning, Data Science, AWS, Python, R, SQL, Tableau) *Pune, MH, IN*

- Designed and deployed a machine learning model using Logistic Regression to predict steam generator failures, resulting in a 40% reduction in maintenance costs and an improved mean time between failures (MTBF) of 20%.
- Collaborated with three members to devise a pump health monitoring solution to plan the preventive maintenance of the pumps leveraging Python for RUL modeling and Tableau for visualization & dashboarding.
- Created automated data cleaning processes using ETL pipelines and efficiently stored the cleaned IoT event data in the OSI PI-historian database, significantly reducing manual processing time by 30%.
- Constructed a Python script to completely automate the process of equipment name extraction from Industrial CAD drawings employing pytesseract and CV2 packages, eliminating manual interventions.
- Managed a team of three graduate engineer trainees, helped set goals and objectives, providing feedback and support.

Academic Projects

ETL Pipeline For Keyword Extraction | *Python, AWS, Java-Script, GitHub Link* **Dec 2022**

- Constructed an ETL pipeline, utilizing AWS Lambda, DynamoDB, and S3 bucket, to extract relevant keywords from multiple websites based on user input which helped boost the SEO score of new websites.

NASA Turbo-Jet Engine Failure Prediction | *Python, Jupyter-Notebook, GitHub Link* **Nov 2022**

- Designed and built a predictive maintenance solution for Jet Engines utilizing Logistic Regression, Support Vector Classifier, and Decision Tree algorithms, preventing expected failures by 20% and improving reliability.

Pneumonia Detection using Transfer Learning | *Python, Pytorch, GitHub Link* **Oct 2022**

- Implemented a pretrained ResNet50 architecture to detect pneumonia in chest x-rays with 85% Recall.

Honors & Awards

- Received Dean's MITA Merit Scholarship at Rutgers Business School.
- Ranked among the top 5% of Data Scientists in the Kaggle Community.
- Elected as the Coordinator of DESCON Hobby Group under the Science & Technology Council, IIT Kanpur.