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: 3.97/4.0*

Indian Institute of Technology, Kanpur **Jul 2015 – May 2019**
Bachelor of Technology in Electrical Engineering *GPA: 3.40/4.0*

Technical Skills

Languages: Python, R, SQL, Cypher (CQL)
Tools: Jupyter Notebook, R Studio, 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

Larsen & Toubro Infotech **Jul 2019 – Sep 2022**
Senior Engineer (Data Scientist) *Bakersfield, CA, US*

- 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.

IDEAS - A SAS Company **May 2019 – Jul 2019**
Data Science Internship *Bloomington, MN, US*

- Utilized the ARIMA model to forecast hotel prices and selected the optimal Auto-Regressive lags using BIC & AIC, resulting in an optimum price point for Boutique Hotels and a subsequent 5% increase in room occupancy.
- Analyzed hotel pricing data, detecting pricing trends and patterns, including seasonal and special event impacts, leading to increased revenue for the hotel chain.
- Improved upon the previous forecasting model by achieving a 10% decrease in sMAPE.

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.
- Created an AWS Lambda function to perform the web-scraping to collect text from websites, applied text cleaning operations, and extracted keywords using YAKE & pretrained KeyBERT models.
- Performed automated write-back of keywords data into DynamoDB and optimized execution time of the ETL pipeline.

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.
- Optimized model performance through hyper-parameter tuning with RandomSearchCV and GridSearchCV.
- Achieved an AUROC score of 0.88 and a recall of 0.78 using Logistic Regression for detecting the failure.

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

- Developed & implemented a pretrained ResNet50 architecture to detect pneumonia in chest x-rays with 85% Recall.
- Enhanced model generalization and training data diversity by implementing data augmentation techniques like rotation, zooming, and cropping.
- Deployed Flask-integrated model on Heroku, enabling faster diagnosis, enhancing accessibility for remote healthcare centers, and reducing testing expenses.

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.