

Jwalit Shah

jwalitshah10@gmail.com | (732) 209-2041 | <https://www.linkedin.com/in/jwalit-shah/> | Availability: (May 2024 Full Time)

[Portfolio](#) | [GitHub](#) | [Tableau Public](#)

WORK EXPERIENCE

Data Science Intern | Crewasis, NY, USA

May 2023 – December 2023

- Collected data from over **17.6M** climate related social media articles to identify relevant trending topics for the year.
- Converted raw CSV data to refined pickle files after URL scraping, encompassing over **200** documents per quarter.
- Preprocessed text to achieve **38%** reduction in text length using Lemmatization, eliminating stop words and punctuation.
- Employed **Topic modeling** and **clustering** techniques to train **LDA** and **LSI** models on corpus of text, yielding **10 clusters per quarter**.
- Presented **5 idea dashboards** using clusters over a **4-week sprint**, facilitating client comprehension and decision-making.
- Enhanced client interaction time up to **45%** by merging **NLP-based keyword** analysis and dynamic **Tableau** visualizations.

Technical Analyst | Paras Engineering Co, India

August 2020 – December 2021

- Built **statistical models of time series data** for inventory demand forecasting, enhancing stock availability to **100%**.
- Estimated Remaining useful life (RUL) of bearings using **predictive modelling** and reduced the maintenance cost by **22%**.
- Performed **ad-hoc analysis** on inventory data using **SQL** to identify optimal inventory levels, leading to a **4%** rise in profit.
- Defined **KPIs** and built interactive **Tableau dashboards** for monitoring Inventory level, Consumption and Lead time.

LEADERSHIP EXPERIENCE

Graduate Teaching Assistant (Data Mining in Engineering) | Northeastern University, Boston, MA

January 2023 – May 2023

- Facilitated Professor's work by conducting weekly Python sessions and grading Homeworks, Quizzes, and Exams.
- Collaborated with a team of 4 TA's and 8 Graders to manage class of 200 students and provide assistance with doubt solving.
- Mentored around 25 students in Academic Projects involving extensive use of Machine Learning concepts.

ACADEMIC PROJECTS

Taxi Demand Forecasting (Skills: Machine Learning, Python, Time Series Forecasting)

January 2023 – April 2023

- Implemented EDA and feature engineering on **30M** NYC yellow cab trips during January 2015 to extract features and eliminate outliers.
- Employed **K-means clustering** to divide New York into equal spatial division and segmented time into 10-minute bins.
- Predicted pickups for cab drivers by **Linear Regression**, **Random Forest** and **XGBoost**, selecting best model with **12.94% MAPE**.

Pneumonia Detection using CNN (Skills: Neural Networks, Deep Learning, Machine Learning)

January 2023 – April 2023

- Preprocessed **5.2k** chest X-ray images, including resizing images to standard size, normalizing pixel values and data augmentation.
- Trained CNN model involving convolution and pooling layers, achieving an accuracy of **91%** in detecting Pneumonia.
- Demonstrated potential of CNN-based approaches in improving efficiency of Pneumonia detection, leading to timely diagnosis and treatment.

Predictive Maintenance Analysis (Skills: Machine Learning, Data Mining, Python)

September 2022 – December 2022

- Conducted **statistical analysis** and **predictive modeling** on a dataset of **10k** records, **14** features to understand reasons for Machine failure.
- Developed model for failure prediction, addressing data imbalance through **SMOTE** and oversampling minority class to **50%**.
- Strategized thorough model evaluation, using precision, recall, and F1-score to identify best performing model with a recall of **0.89**.

Application for Blood Bank Management System (Skills: MySQL, MongoDB, Python, Tableau)

September 2022 – December 2022

- Developed an application for Blood Bank Management System to simplify and automate the process of searching blood.
- Designed **MySQL Database** and populated it for real-world simulation, enabling **CRUD operations** on 9 tables.
- Incorporated table level check constraints, triggers, stored procedures using advanced SQL queries and Python-based visualizations.

EDUCATION

Master of Science in Industrial Engineering (GPA: 3.95/4.0) (Northeastern University - Boston, MA)

January 2022 – May 2024

Relevant Courses: Machine Learning, Natural Language Processing, Data Visualization, Probability, Statistics, Database Management

Bachelor of Technology in Mechanical Engineering (IITRAM – India)

August 2016 – June 2020

Relevant Courses: Python Programming, Engineering Mathematics, Operations Research & Project Management

TECHNICAL SKILLS & CERTIFICATIONS

Programming Languages: Python (Pandas, NumPy, Matplotlib, Seaborn, TensorFlow, Keras), SQL, R, HTML, CSS, Bootstrap

Analytical Tools: Tableau, MS Office, MS Excel, Jupyter Notebook, RStudio, Power BI, MS Outlook, Google Data Studio

Databases: Google Cloud Platform, MySQL, MongoDB, Apache Spark, Neo4J, Amazon Redshift

Machine Learning: Regression, Classification, Clustering, PCA, Hypothesis Testing, Data Analysis, Decision Modeling