Badreesh Shetty

617-971-7369 | linkedin.com/in/badreesh-shetty | badreeshshetty@gmail.com | badreeshshetty.github.io

WORK EXPERIENCE

Centelon   **Mumbai, India**
Data Scientist   01/2020 - 05/2022

- Executed an AWS Python-based ETL pipeline for a real-time OCR tool, handling **30,000+** financial documents (**100 GB+**) from email attachments for a leading pharmaceutical client. Realized a **48%** decrease in manual effort, resulting in a monthly time savings of **50 hours**.
- Led a team of **3** data scientists in conceptualization and execution of a fuzzy string-matching algorithm that improved accuracy and speed of key information extraction resulting in a **17%** decrease in processing time.
- Built the backend infrastructure using Django REST API framework and SQL which facilitated client's access to a CRUD web-app functionality through monthly reports, delivering valuable analytics for data-driven decision-making.
- Utilized Plotly-Dash and Excel to create interactive visualizations and dashboards, improving stakeholders' comprehension of key findings.
- Collaborated with cross-functional teams to maintain the codebase and established streamlined continuous integration using Git for efficient project development and version control.

AAIC Technologies   **Mumbai, India**
Machine Learning Intern   12/2018 - 04/2019

- Designed Sentiment Analysis for Amazon Fine Food Reviews using SVM, K-Nearest Neighbors, and ensemble methods (Random Forest, Gradient Boosting), performed GridSearchCV to find the optimized model with **0.96** F1 score in Gradient Boosting.
- Performed classification machine learning models for Human Activity recognition using Logistic Regression, Decision Tree, LSTM, and Random Forest resulted in an **83%** accuracy on test data.

ACADEMIC PROJECTS

Demand Sales Forecasting for Thermofisher (Capstone)

- Forecasted sales demand for Thermos Fisher's Biosciences division using a hybrid approach, integrating regression, time series, and deep learning models, yielding a **18%** MAPE.
- Integrated flu patterns, stock market data, healthcare ETF stocks, and sentiment data from Google News as feature engineering, enhancing the predictive capabilities and overall robustness of the developed models.

Recommender System for Student Housing in Boston, Website and Github

- Created a geospatial web app and a Tableau dashboard, utilizing scraped data from Zillow and Analyze Boston, simplifying property and event searches for Boston students by zipcode.

SKILLS & AWARDS

- **Languages, Frameworks, & Databases:** Python, R, Django, Flask, FastAPI, PostgreSQL, MongoDB, Redis
- **Tools & Libraries:** Snowflake, AWS, Alteryx, Databricks, Airflow, Kafka, Cassandra, Airbyte, Redshift, BigQuery, Hive, PowerBI, Pandas, Keras, Scikit-Learn, Pytorch, Pyspark, Matplotlib, Seaborn, Docker, Streamlit, DBT, Cron
- **Machine Learning Algorithms:** PCA, Kmeans Clustering, Naïve Bayes, Decision Tree, Xgboost, CNN, RNN
- Awarded **1st Prize** Bhaskaracharya Scholarship in a stock alpha prediction Hackathon with 30+ participants.
- Secured **3rd Prize** in Technical Blog Competition @AppliedAICourse for Data Visualization using Matplotlib.
- Received "SPOT Award" from the Director of Data Science in recognition for outstanding performance.

PUBLICATIONS

Garnered **1 million+ views** and **50+ citations** @ Google Scholar with focus on machine learning and data science.

- Syndicated **4** articles in Built-In, a reputable publisher as an expert contributor in the Data Science Domain.
- Authored and contributed articles on Medium's TowardsDataScience platform, gaining 1,100 followers.

EDUCATION

Northeastern University   **Boston, MA**
Master of Science in Business Analytics   09/2022 - 12/2023
Graduate Research Assistant, Marketing Department

- Teamed with Professor Jay Mulki to curate and assess marketing articles and assignments for **50+ students**.
- Compiled and analyzed research papers on "professionalism in sales" to help with the Professor's thesis focus.

International School of Engineering   **Mumbai, India**
Post Graduate Program in Data Science   06/2019 - 11/2019

University of Mumbai   **Mumbai, India**
Bachelor of Technology in Computer Science   06/2014 - 08/2018