

Shreyas Desai

Software Developer/Machine Learning Engineer/Data Engineer
portfolio — shreyasdesai3013@gmail.com — +1(201)-492-8801

LinkedIn: shreyas-desai

GitHub: shreyas-desai

Education

- * **Master's in Computer Science, 3.67**
- * **B.Tech. in Computer Science, 8.5**
- * **Diploma in IT, 81.53 %**

Stevens Institute of Technology, September 2023 - May 2025
Department of Technology, SUK, September 2019 - July 2022
BVIT, Navi Mumbai, August 2016 - May 2019

Experience

* Software Engineer

Extrapreneurs India Pvt.Ltd, November 2021 - July 2023

- Developed functional **FastAPIs** for services, including Market Data collection, User Data management, and Bulk Data Load.
- Maintained legacy **Java APIs** based on **SpringBoot** framework, incorporating updates for globalization and localization throughout the system.
- Integrated and synchronized data between **PostgreSQL** and **Salesforce**, enhancing data accuracy and accessibility. Utilized **SQL** for database operations.
- Utilized **Salesforce Bulk Data API** to directly load data from a third-party site to Salesforce Bulk Object, with **AWS S3 Bucket** backup.
- Implemented successful data transfer to custom Salesforce objects using **AWS serverless lambdas** and **AppFlows**.
- Experience with **Git CLI** and **Desktop** for code version control.

* Data Engineering Intern

SessionAI (previously Zineone), August 2022 - February 2023

- Provided recommendations of offers based on retrieved user session activity using **LSTM models** with hyperparameter tuning and visualizing results for improved feature selection.
- Managed **data manipulation** and **feature extraction**. Implemented **Data Visualization** for feature selection
- Performed collaborative experiments on **Google COLAB** and **Jupyter** notebooks and trained multiple **ML regression models** for session-based user data.
- Evaluated inferences from the resulting metrics and deployed models on **MLFlow** for testing and production.
- Analyzed data from over 60000 monthly active user sessions, showcasing expertise in **Data Analysis** and **Statistical Analysis**.

Projects

* Income Classify

Unveiling Socio-Economic Thresholds, Project

- The project employs data mining and knowledge discovery techniques to classify individuals into salary brackets ($\leq 50k$ and $> 50k$) based on demographic features.
- Addressing income disparity, the analysis focuses on age, gender, ethnicity, and education, aiming to provide insights into fair pay practices.
- The predictive model utilizes a comprehensive dataset to determine the reliability of these features in classifying individuals' income levels (0 for $\leq 50k$, 1 for $> 50k$).

* PhishURL

URL Security: Phishing Detection Insights, Project

- The "PhishURL" project involved building a dataset of 55
- Through rigorous training of machine learning models such as Random Forest, SVMs, Decision Trees and Bayesian Networks, accuracy rates ranging from 92
- To ensure practicality, a user-friendly web application was developed using Flask that provides immediate phishing risk evaluation for submitted URLs was developed.

* Petals to the Metal

Kaggle Competition, Code

- Applied data augmentation techniques to provide model with a diversified data during training to avoid overfitting.
- Trained data on ensemble as well as fine tuned standalone models to check improvements in accuracy.
- Made use of MobileNetV2 pre-trained model with 'imagenet' weights over a customised sequential model for best accuracy (experimented with MobileNet, DenseNet201, VGG19).

* Natural Language Processing with Disaster Tweets

Kaggle Competition, Code

- Applied appropriate data preprocessing strategies to make the data trainable.
- Trained the data on distilBERT, a variant of the popular BERT model
- Trained the data on BERT preset model and BERT preprocessor with a total of 66,955,010 trainable parameters to achieve an f1-score of 0.9 on training data and 0.77 on validation data.

Skills

- **Programming Languages:** Python, Java, SQL, C++
- **Software Proficiency:** VS Code, Eclipse IDE, WSL, PyCharm, Anaconda
- **Data Analysis:** Power BI, Feature Extraction, Data Visualization (MatLAB, Matplotlib), Statistical Analysis, Exploratory Data Analysis (EDA), Data Wrangling, Collaboration Tools(Google COLAB, Jupyter)
- **Decision Sciences:** Quantitative Analysis, Decision Trees and Decision Analysis, Predictive Modeling
- **Backend Frameworks:** FastAPI, Flask, Django, SpringBoot
- **Serverless Technologies:** AWS Lambda, GCP Functions