# Shreyas Desai

*Software Developer/Machine Learning Engineer/Data Engineer*
portfolio — shreyasdesai3013@gmail.com — +1(201)-492-8801
**LinkedIn:** shreyas-desai **GitHub:** shreyas-desai

*Experienced Software Developer, Machine Learning and Data Engineer with 2 years of industry exposure demonstrating expertise in developing efficient and cost-effective APIs, integrating and synchronizing data and conducting collaborative machine and deep learning experiments. Proven skills in data analysis, feature extraction and model training, seeking to apply a diverse skill set and technical proficiency to contribute to innovative projects in a dynamic environment.*

## Education

**Master's in Computer Science, 3.67** — *Stevens Institute of Technology, September 2023 - May 2025*
**B.Tech. in Computer Science, 8.5** — *Department of Technology, SUK, September 2019 - July 2022*
**Diploma in IT, 81.53 %** — *BVIT, Navi Mumbai, August 2016 - May 2019*

## 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)
- **ML and Deep Learning:** Tensorflow, Scikit-learn, Pandas, Numpy, Decision Trees and Decision Analysis, Predictive Modeling
- **Backend Frameworks:** FastAPI, Flask, Django, SpringBoot
- **Serverless Technologies:** AWS Lambda, GCP Functions

## Experience

**Software Engineer** — *Extrapreneurs India Pvt.Ltd, November 2021 - July 2023*

- Efficiently designed functional **FastAPIs** for services, including Market Data collection, User Data management and Bulk Data Load by revamping the APIs and making it more cost-effective and reducing the response time by **25%**. Modernized legacy Java APIs based on **SpringBoot** framework by incorporating updates for globalization and localization throughout the system.
- Achieved seamless integration and synchronization of data between **PostgreSQL** and **Salesforce**, significantly enhancing data accuracy and accessibility. Employed advanced **SQL** techniques for precise database operations. Successfully utilized **Salesforce Bulk Data API** to orchestrate direct data loading from a third-party site to Salesforce Bulk Object, ensuring a robust AWS **S3** Bucket backup strategy.

**Machine Learning/Data Engineering Intern** — *SessionAI (previously Zineone), August 2022 - February 2023*

- Procured impactful results by delivering targeted recommendations of offers, leveraging **LSTM** models with meticulous hyperparameter tuning. Employed advanced visualization techniques to enhance feature selection, contributing to a more refined and effective user session activity analysis.
- Conducted collaborative experiments on **Google COLAB** and **Jupyter** notebooks, leading to the successful training of numerous deep learning regression models tailored for session-based user data. Evaluated insights derived from resulting metrics, culminating in the strategic deployment of models on **MLFlow** for rigorous testing and merging into production.

## Projects

**Income Classify** — *Unveiling Socio-Economic Thresholds, Project*

- Successfully applied cutting-edge data mining and knowledge discovery techniques in a project aimed at classifying individuals into distinct salary brackets ($<=$50k and $>$50k) based on critical demographic features.
- The analysis, targeting age, gender, ethnicity, and education, strategically addresses income disparity and contributes valuable insights to promote fair pay practices. Implemented an advanced predictive model using a comprehensive dataset, with an accuracy of **94%**, accurately classifying individuals' income levels (0 for $<=$50k and 1 for $>$50k) based on demographic features.

**PhishURL** — *URL Security: Phishing Detection Insights, Project*

- Lead the "PhishURL" project, meticulously curating a dataset comprising 55% benign and 45% malign URLs to create a robust foundation. Achieved remarkable accuracy rates ranging from 92% to 96% through intensive training of machine learning models, including Random Forest, SVMs, Decision Trees, and Bayesian Networks.
- Translated project success into practicality by developing a user-friendly web application using Flask, enabling instantaneous phishing risk evaluation for submitted URLs.

**Petals to the Metal** — *Kaggle Competition, Code*

- Implemented data augmentation techniques to furnish the model with diverse training data, mitigating the risk of overfitting. Conducted comprehensive training on ensemble models and fine-tuned standalone models, systematically assessing and enhancing accuracy.
- Leveraged the MobileNetV2 pre-trained model with 'imagenet' weights, surpassing accuracy benchmarks set by MobileNet, DenseNet201, and VGG19

**Natural Language Processing with Disaster Tweets** — *Kaggle Competition, Code*

- Implemented effective data preprocessing strategies to optimize the data for training. Conducted training on distilBERT, a variant of the renowned BERT model.
- Skillfully trained the data on a BERT preset model and BERT preprocessor, featuring a total of 66,955,010 trainable parameters, resulting in an impressive f1-score of 0.9 on the training data and 0.77 on the validation data