

Satya Shah

Mail-satyashah26102001@gmail.com | M:(+1)6824039713 | [LinkedIn](#) | [GitHub](#)

INTERNSHIPS:

-Computer Vision Intern, Avidclan Technologies, Ahmedabad, India [\[link\]](#)

May 2021 –August 2021

- Developed a web app for CT-scan-based kidney tumor detection for a hospital using **transfer learning** (VGG16 with **Tensorflow/Keras**), achieving **98% recall** and assisting doctors in the early diagnosis of kidney cancer for **500+ patients**.
- Built a complete ML pipeline from data ingestion to model evaluation using version control tools like **DVC**, **Dagshub**, and **MLFlow**.
- Preprocessed the images using Otsu's Binarization, Watershed segmentation, and Normalization. Handled class imbalance by **data augmentation**.
- Developed a user interface using **HTML**, **CSS**, **JavaScript**, **Flask** for routing, and **MongoDB** for storing patient records.
- Containerized the entire application using **Docker**. Deployed the app on **AWS** using **CI/CD** with **GitHub Actions**, **ECR**, and **EC2**.

-Machine Learning Intern, Sensegood Instruments Private Limited, Ahmedabad, India [\[link\]](#)

June 2022 –July 2022

- Built a **predictive analytics solution** that reduced pneumatic braking inspection costs by 20% for truck owners.
- Analyzed **36K+** sensor data points (**175 features**) to detect failures.
- Optimized a **LightGBM** model using **Optuna**, achieving **80% recall** and **88% accuracy** through advanced preprocessing techniques, including **KNN Imputer**, **SMOTE-Tomek**, and **KS test**.
- Deployed an **end-to-end ML pipeline** from data ingestion to model evaluation with **Fast API**, **MongoDB**, **AWS (S3, EC2, ECR)** and **GitHub Actions**.

-Computer Vision Intern, Technostacks Infotech Private Limited, Ahmedabad, India [\[link\]](#)

January 2023 –May 2023

- Developed an automated invoice parsing web app using **OCR**, extracting data with **95% accuracy** and enabling **CSV exports** for manufacturers.
- Built with **Python**, **Flask**, **HTML**, **CSS**, **JavaScript**, integrating **MongoDB** and **OpenCV** for OCR processing.
- Implemented **regex-based parsing** for structured invoice data, allowing **updates, additions, and deletions**.

PROJECTS:

-Smart Fail- Machine Failure Predictor [\[link\]](#)

January 2025

- Predicted machine failures with a **macro F1 score of 86%** using **Logistic Regression** on **100K+ rows** after thorough preprocessing and EDA.
- Dockerized** the application, implemented **Jenkins** for continuous integration, and **Argo CD** for continuous deployment.
- Deployed** on a **GCP VM instance** using **Minikube** for efficient **Kubernetes-based** deployment.

-Hotel Reservation Prediction [\[link\]](#)

March 2025

- Developed a full-stack machine learning application to predict hotel reservation cancellations with **91% precision**, enabling proactive measures that could reduce revenue loss by up to **15% due to no-shows and last-minute cancellations**.
- Performed end-to-end data pipeline including preprocessing, EDA, handling class imbalance, and training a **LightGBM** model; visualized insights using **Tableau dashboards**.
- Dockerized** the app, used **Jenkins** for **CI/CD**, deployed on **GCP**, with frontend in **HTML/CSS**.

-Tableau Dashboards: User Behavior Analysis [\[link1\]](#) & Stock Market Analysis [\[link2\]](#)

April 2025

- Built dynamic **Tableau** dashboards using datasets from platforms like **Netflix**, **Amazon**, **Goodreads**, and **Airbnb**.
- Developed an interactive dashboard to analyse **stock data** of **Apple**, **Google**, **Netflix**, **X**, **Tesla**, and **META** from 2019 to 2024.
- Created **custom features** and utilized **dynamic charts and KPIs** to enhance visualization and insights for an enhanced user experience.

TECHNICAL SKILLS:

Data Science/Machine Learning/ Deep Learning/ Computer Vision/NLP – Supervised Learning algorithms, Unsupervised Learning algorithms, statistics, Exploratory Data Analysis, ANN, CNN, RNN, Transformers

Programming languages - Python, HTML, CSS, SQL

API/MLOPS- Flask, Fast API, GitHub, GitHub actions, Docker, Kubernetes, MLFlow, DVC, Dagshub, Jenkins(CI/CD), Argo CD

Databases and Cloud- MySQL, MongoDB, AWS(S3,ECR,EC2), GCP

Data Visualization- Tableau, Excel

Python packages and Libraries- Scikit-Learn, TensorFlow, Keras, PyTorch, NumPy, Pandas, SciPy, Seaborn, Matplotlib, OpenCV, Optuna

EDUCATION:

-Bachelor of Technology (BTech), Computer Science,

August 2019- May 2023

School of Engineering and Applied Science, Ahmedabad University, Gujarat, India

-Masters of Science(MS), Computer Science,

August 2023- May 2025

University of Texas, Arlington, Texas, United States of America, GPA (3.9/4.0)

-**Soft Skills:** Strong problem-solving skills, excellent written and verbal communication skills, adaptability, Teamwork, critical thinking

-**Kaggle Competition [\[link\]](#)**- Participated in a Kaggle competition on mushroom edibility prediction, leveraging machine learning for binary classification of **1 million rows**. Achieved a top 25% leaderboard ranking among 2500 participants.

-**UTA'25 Datathon Winner [\[link\]](#)**- Build a RAG-based application for pediatricians offering instant and medical insights to support clinical decisions with an average response time under 3 seconds, winning among 50 teams that participated.