# Soham Gore

Mumbai, India | acad.soham@gmail.com | linkedin.com/in/sohamgore | github.com/debug-soham

## Professional Description

Enthusiastic engineering student with a strong foundation in machine learning, statistics, and data analysis. Skilled in building and evaluating models using Python and Scikit-learn, with hands-on experience in data preprocessing, feature engineering, and model deployment fundamentals.

## Technical Skills

**Programming & Tools**: Python, NumPy, Pandas, Scikit-learn, SQLite3, Java, C, Git, Docker

**Data Analysis**: EDA, Statistics, Probability, Hypothesis Testing, Data Visualization, Probability Distributions

**Machine Learning**: Linear & Logistic Regression, Decision Trees, Random Forest, Gradient Boosting, XGBoost, SVM, Naïve Bayes, KNN, Clustering, PCA, Anomaly Detection

**Feature Engineering & Optimization**: Feature Selection, Data Scaling, Encoding, Regularization, Hyperparameter Tuning, Cross-Validation

**Deployment**: Flask, Streamlit, Docker (Basics)

## Experience/Internships

**Open Source Contributor** **October 2025 - Present**
*Hacktoberfest*

**Open Source Contributor** **July 2025 - Present**
*GSSoC - GirlScript Summer of Code*

- Contributed to the development of 'etsi-watchdog', a Python library for real-time data drift detection in machine learning pipelines, by implementing new statistical algorithms and enhancing monitoring capabilities.

**Technical Team Member** **July 2025 - Present**
*Datazen Somaiya* *Mumbai - On-site*

- Collaborating on data-centric community projects focused on AI literacy and ethical ML.
- Leading events and hackathons as part of the technical team.

**AI & ML Intern** **June 2025 - July 2025**
*Elevate Labs* *Virtual*

- Engineered a Convolutional Neural Network (CNN) in Keras/TensorFlow that achieved 75% accuracy in classifying 10 distinct music genres from the 1,000-track GTZAN dataset by analyzing Mel-Frequency Cepstral Coefficients.

## Projects

**Music Genre Classification** | *Deep Learning - TensorFlow, Keras, Librosa, NumPy*

- Developed a Convolutional Neural Network (CNN) achieving approx 75% accuracy on the GTZAN dataset for classifying 10 music genres.
- Extracted Mel-Frequency Cepstral Coefficients (MFCCs) as key audio features to train and validate the model.
- Implemented efficient data preprocessing and model optimization for balanced genre recognition.

**Personal Finance Risk Classifier** | *Machine Learning - Python, Scikit-learn, Pandas, Seaborn, CLI*

- Built a Gradient Boosting model to classify users as Conservative or Aggressive investors based on financial data.
- Engineered risk profiles using demographic and behavioural metrics, integrated into an interactive CLI tool.

## Education

**KJ Somaiya School of Engineering, Mumbai** **2024 - Expected 2028**
*Bachelor of Technology in Electronics and Computer Engineering* *CGPA: **8.95** / 10 (After 1st Year)*
*Honours in Data Science and Analytics*

## Relevant Courses & Certifications

**Machine Learning Specialization** **June 2025**
*DeepLearning.AI, Stanford Online*