· Hello World! ·

Harsh Singhal

Data Scientist | Goldman Sachs | IIT Roorkee

harshsinghal726@gmail.com
+91-7999025413

> Give me the data, and I will turn it into a story
> Skilled in developing machine learning models and gaining actionable insights from data through statistical analysis in order to make sound business decisions.

Education

Indian Institute of Technology, Roorkee — Roorkee, India
Bachelor of Technology; CGPA: 7.7 — May 2018

Experience

Associate | Goldman Sachs — Jan 2021 - Present

Used settlement data to predict the probability of failure of any trade

- Built model using CatBoost and inspected SHAP values to generate commentary on top reasons for trade failure; created a Plotly dashboard for backtesting
- Used SMOTE to handle imbalanced classes along with K-fold target encoding of high-cardinality categorical variables to achieve an 0.9 AUC, 60% precision, and 58% recall

Semi-automated the matching process of inbound payments

- Developed a multi-class random forest model to improve the matching process of inbound payments, saving $200M per year
- Used TFIDF method to extract features from raw text data to achieve an accuracy of 80%.

Analyst | Goldman Sachs — June 2018 - Dec 2020

Systematically captured and explained drivers of Unencumbered Securities worth $15B

- Created attributes for clustering like stickiness, persistence among other behavioral features
- Explored K-means, DBSCAN and hierarchical clustering; used silhouette score and CH-index to optimize the number of clusters

Built a tool to automate the Exploratory Data Analysis (EDA) process

- Key features include descriptive statistics, variable associations, target variable characteristics, basic data quality checks, and missing value analysis

Data Scientist | Datametica Solutions Private Limited — May 2017 - July 2017

Developed a framework for Optical Character Recognition (OCR) of a newspaper

- Created a digitization workflow that segments the entire newspaper image at the article level and extracts the text from each segment.

Offline Signature Verification using Deep Convolutional Neural Network (CNN)

- Built a model on top of the VGG16 architecture and trained it with transfer learning on the ICDAR SigComp dataset to achieve a 70% accuracy and a 14% false acceptance rate

Scholastic Achievements

- Secured **1st position** in GS Quantify 2017 competition by Goldman Sachs for 22 top-tier colleges
- Secured **3rd position** in Analyze This 2017 competition by American Express (1k+ teams)

Personal Projects

- **WhatsApp chat analyzer**: In-depth exploratory data analysis of WhatsApp group chat
- **Nifty 50 Performance in the last decade**: Analysis of annual returns of widely tracked index