

# Shreyas Kulkarni

## Data Scientist

✉ shreyas.kulkarni086@gmail.com

☎ 8605370416   **in** LinkedIn   **📖** Blog

### SUMMERY

With 2.5+ years of experience, I am well-versed in Python, Machine Learning, Deep Learning, NLP, and Computer Vision related use cases. I have exposure to the entire model development life cycle, which includes data preprocessing, EDA, model development, deployment, and evaluation.

### SKILLS

Python, Data Preprocessing, Data Integration, Data Cleaning, Exploratory Data Analysis, Model Development, Machine Learning, Deep Learning, Natural Language Processing (NLP), Computer Vision, Git, GitHub, Jupyter Notebook, Google Colab, Visual Studio Code (VSCode)

### EDUCATION

**Bachelor's in Engineering, CGPA 7.32, Pune University**  
2017 – 2020 | Pune

### CERTIFICATES

- Machine Learning Specialization -- (DeepLearning.AI)
- Data Science Training (6 months Duration) -- (ATI, Bangalore)
- Statistics -- (365 DS)
- Crash Course on Python -- (Google)
- Introduction to Data Analytics -- (IBM)
- From SQL Basics To MySQL Mastery (Udemy)


### AWARDS

#### **Data Gladiator, Indium Software**

Recognized for exceptional performance, receiving client appreciation,

### WRITING PORTFOLIO

**Responsible AI, Indium Software** 

**Personal Blogs, Shreyas Kulkarni** 

### PROFESSIONAL EXPERIENCE

#### **Junior Data Scientist, Indium Software**

November 2021 – Present | Bangalore

#### **Associate Intern, Splashgain Technology Solutions**

December 2020 – July 2021 | Pune

### PROJECTS

#### **Healthcare/Insurance Data Extraction Pipeline,**

**Client: Policy Reporter**

**Tools Used:** Pandas, NumPy, TensorFlow, Transformers, spaCy, Detectron2, OpenCV, Git, GitHub, pytest, Camelot, Tabula, LabelMe.

- Developed models based on Keras CNN and Huggingface ViT Transformer for classifying various table types and accurately identifying border types, and custom models for document categorization based on its unique types.
- Built models to predict column lines, using Detectron2 utilizing annotation tools and image processing.
- Contributed to developing NLP model using Spacy to extract specific elements from drug names.
- Developed automated pipeline for processing 50+ templates in healthcare/insurance domain, used primarily Camelot and Tabula for table data extraction, used Git for code delivery, and managing and addressing changes requested by clients.

#### **Business-to-Business Networking Enhancement,**

**Client: BNI Global**

**Tools Used:** Pandas, Modin, Scikit-Learn, TfidfVectorizer, Word2Vec.

- Performed data preprocessing and EDA to extract meaningful insights.
- Experimented with content-based and collaborative filtering methods.
- Developed recommendation models utilizing TfidfVectorizer, Word2Vec, KNN, vectorization techniques, and cosine similarity.

#### **Data Integration and Enhancement Project,**

**Client: Higher Education Institutions and Colleges**

**Tools Used:** Python, Pandas, Microsoft Excel, Google Sheets

- Performed data pre-processing and cleaning across different data sources to make them suitable for the product's template format.
- Developed Python scripts to handle common data source issues.
- Collaborated with multiple domestic and international clients, including Pune University, Mumbai University, KC and HR Mumbai, IIIT, and IIM, Credforce, WeSchool

### VOLUNTEERING

#### **BPB Publications, Technical Reviewer**

May 2023 – Present

Contribution to improving the technical accuracy of the book's content.

#### **Data For Good, Contributor/ Volunteer**

2021 | Canada

Contributed to Social media platforms to increase it's reach.