

# Shreyas Kulkarni Data Scientist

✉ shreyas.kulkarni086@gmail.com 📞 8605370416 [in LinkedIn](#) 🌐 helloshreyas.com

📍 Pune, Maharashtra, India (Open to relocate)

## 📄 SUMMARY

---

Data Scientist with 3 years of hands-on experience in the field of data science. specializing in working with Python, Machine Learning, and Deep Learning. skilled in developing advanced models and performing data processing and analysis.

## 📁 PROFESSIONAL EXPERIENCE

---

### Junior Data Scientist

Nov 2021 – Present

*Indium Software*

#### Project: Automated Formulary Parser for Healthcare Documents

Tech Stack: Python, Pandas, NumPy, TensorFlow, Transformers, spaCy, Detectron2, OpenCV, Git, GitHub, pytest, Camelot, Tabula, LabelMe.

- Built an Image Classification models to differentiate between types of 15+ templates, enabling separate extraction pipelines for these files, covering 40% of templates.
- Developed models to predict column lines for templates missing gridlines, covering 25% of templates, using Detectron2.
- Developed NLP model using spaCy to accurately identify and separate dosage entities, as a part of post-processing of all types of templates.
- Developed end-to-end automated pipeline including integration of all models as well as integration of extraction code.
- Implemented test cases using pytest to ensure the coverage and reliability of a project.

#### Project: People-to-People Recommendation Engine

Tech Stack: Pandas, Scikit-Learn, TfidfVectorizer, Word2Vec, KNN.

- Developed a demo of a recommendation system using content-based techniques like TfidfVectorizer and Word2Vec.
- Built an end-to-end pipeline and presented a demo to global C-level executives, showcasing business value and potential impact.

### Associate

Dec 2020 – Jul 2021 | 8 Months

*Splashgain Technology Solutions Pvt. Ltd.*

- Improved data pre-processing and importation time by 70% by developing automated scripts for data preprocessing and cleaning through the use of Python, and pandas.
- Collaborated with clients to identify recurring issues, and formulated SOPs to eliminate confusion and repetitive workarounds enhancing client efficiency by 30%.

## 🎓 EDUCATION

---

### Data Science Course (Analytics 360)

Jul 2021 – Feb 2022 | 8 months

*Analytics Training Institute from Redwood Algorithms* [🔗](#)

### Bachelor's in Engineering (First Class)

2017 – 2020

*Pune University*

## 🧠 SKILLS

---

### Primary Skills


Python (NumPy, Pandas, Scikit-learn, Pytorch, Keras, pytest), Random Forest, K-Means, KNN, CNN, Vision Transformer (ViT), Detectron2.

### Tools

Jupyter Notebook, Jupyter Lab, Git, GitHub, Google Colab, VS Code.

## CERTIFICATIONS

---

- **AWS Certified Cloud Practitioner, AWS** 
- Machine Learning Specialization, DeepLearning.AI/Coursera
- LLMs Mastery: Complete Guide to Transformers & Generative AI, Udemy
- Statistics, 365 Data Science
- Crash Course on Python, Google
- Introduction to Data Analytics, IBM
- From SQL Basics To MySQL Mastery, Udemy
- Digital Product Management: Modern Fundamentals, Coursera

## INTERESTS

---

blogging, technical writing, investing and finance