

# Harshal Honde

Pune, IN • 9021664306 • harshalhonde50@gmail.com • linkedin.com/in/harshal-honde268 •  
https://github.com/Harry262000

## WORK EXPERIENCE

**Data Science Intern** • Pune , Maharashtra  
Nroad

May 2024 - Present

- Executed full OCR workflow for annual financial reports of clients using PyTesseract, PyMuPDF including detection and conversion of scanned pages to editable text, seamlessly integrating with subsequent data pipelines.
- Designed and deployed NER models leveraging PyTorch and TensorFlow to extract and tag entities from PDF documents, significantly enhancing data extraction accuracy and enabling effective data utilization for LLM models .
- Accelerated document processing operations through automation of OCR and NER tasks, significantly reducing manual processing time and enhancing the overall efficiency of financial report analysis.

## PROJECTS

### Advanced NLP-Powered Text Summarization Tool

- Leveraged **LSTM** and **BERT** models to develop an advanced text summarization tool, enhancing data processing efficiency.
- Utilized **PyTorch**, Hugging Face **transformers**, and Streamlit to create an interactive application for extractive, abstractive, and hybrid summarizations, significantly improving the user experience by providing quick and relevant information.

### Interactive PDF Analyzer

- Processed and analyzed **unstructured data** from financial PDF documents, creating a tool that extracts and visualizes tables and text using **NLP** and data extraction algorithms to convert complex financial information into actionable insights.
- Integrated **fine-tuned** Hugging Face models with libraries (**pdfplumber**, **camelot**) to enhance **table detection** and **extraction** from PDFs. Applied a **fallback approach** to handle both **structured** and **unstructured data**, ensuring accurate and comprehensive **information retrieval**.

### ConversationFilterAI

- Designed a real-time **voice conversation** filtering system to monitor and address abusive language during live **customer-executive calls**.
- Leveraged Hugging Face **Transformers** for NLP models, enabling precise transcription and analysis of voice data in real-time.
- Implemented a **deep learning speech-to-text pipeline** using **PyTorch**, converting live audio streams into text for instant detection of abusive language.
- Utilized **NLP techniques** for language understanding and filtering, ensuring the system effectively identifies and flags inappropriate language during conversations.

### CoinMarketCap Cryptocurrency Data Scraping

- Extracted cryptocurrency market capitalization data from CoinMarketCap using web scraping tools like **requests** and **pandas**.
- Analyzed the performance of various cryptocurrencies to identify trends crucial for making informed investment decisions.
- Utilized insights from the analysis to inform investment strategies and capitalize on emerging trends in the cryptocurrency market.

## SKILLS

**Hard skills:** AWS, Deep Learning, Machine Learning, NLP, Python, Pytorch, scikit-Learn, SQL, TensorFlow, Time Series Forecasting

**Soft Skills:** Business Acumen, Communication, Critical Thinking, Product Understanding, Storytelling

## EDUCATION

**BE in CS & Electrical engineering**  
**Pune University** • GPA: 8.3

Aug 2019 - Dec 2023