Harshal Honde

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

Data Science Intern

Highly motivated Entry level Machine Learning Engineer with a passion for developing innovative data processing solutions. Skilled at designing and implementing machine learning models, optimizing data extraction pipelines, and improving system efficiency. Proven ability to enhance data processing workflows and contribute to cross-functional projects.

WORK EXPERIENCE

Data Science Intern • Pune , Maharashtra   May 2024 - Present
nroad

- Designed and implemented a robust PDF processing pipeline to convert PDF files into images and extract text for analysis and storage.
- Automated text extraction and image saving from PDFs, ensuring efficient handling of large documents with error handling and resource management.
- Extracted and analyzed text data from PDFs, utilizing techniques to filter out highlighted and struck-out content.
- Managed file system operations by organizing image and text output directories and ensuring compatibility across different file formats and structures.

PROJECTS

PDF Page Classifier

- Developed a Python tool to classify PDF pages as scanned, text, or partially scanned, utilizing text length and image area thresholds for accurate classification.
- Created a Flask web application that allows users to upload PDFs and view detailed analysis results, leveraging PyMuPDF for efficient processing and precise results.
- Published the project on GitHub, providing comprehensive documentation, usage instructions, and examples to facilitate user understanding and encourage contributions.

Web Scraping with Scrapy and Playwright

- Utilized Playwright to handle dynamic web content, enabling interaction with JavaScript-rendered elements.
- Developed a Scrapy spider to efficiently retrieve and parse structured JSON data from API endpoints.
- Ensured full compatibility and seamless operation on Linux environments for both web scraping and API data retrieval tasks.

Real Time Chat Application

- Developed a real-time chat application using Flask and Socket.IO, enabling seamless communication between multiple users in a web-based environment.
- Implemented WebSocket communication protocols to facilitate instant messaging and real-time updates, enhancing user engagement and interactivity.
- Utilized Flask-SocketIO library to integrate WebSocket functionality with Flask application, ensuring smooth bidirectional communication for chat message exchange.
- Utilized AWS services such as EC2 for virtual server hosting and application deployment to deploy the real-time chat application on a scalable and reliable cloud infrastructure.

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.

Time Series Electricity Demand Forecasting

- Crafted a cutting-edge time series forecasting model for electricity demand prediction.
- Leveraged advanced statistical techniques and machine learning algorithms in Python to accurately forecast future electricity demand patterns.
- Recognized for the model's accuracy and reliability, earning acclaim from professors and peers for its potential to inform strategic decision-making and resource allocation in energy management.

SKILLS

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

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

EDUCATION

BE in Electrical engineering
Pune University • GPA: 8.3   Jan 2019 - Dec 2023

CERTIFICATIONS

Machine Learning Practitioner   Mar 2022 - Aug 2023
DeepLearning.AI

Data Science Professional Certification   Apr 2022 - Jan 2023
IBM