RUPANSHU KAPOOR

Jaipur, India • +91-9983810448 • rupanshukapoor@outlook.com
Portfolio GitHub LinkedIn

SKILLS

Data Analysis: Python (Pandas, NumPy, SciPy), Matplotlib, Seaborn, MySQL, Tableau, PowerBI
Machine Learning: scikit-learn, TensorFlow, Keras, PyTorch, ANN, CNN, NLP, OpenCV, Statistical Analysis
Generative AI: LangChain, HugginFace, LLM, VectorDB, Pinecone, ChromaDB, GAN
Big Data Technologies: Hadoop, Apache Spark
Additional Tools: Git, DVC, Docker, JIRA, AWS, GCP, Flask

EDUCATION

Post Graduation Program- Data Science and Machine Learning **Nov 2023 - Aug 2024**
Imarticus Learning

Bachelor of Engineering - Electronics and Computer **Aug 2015 - July 2019**
M.B.M Engineering College, Jodhpur

WORK EXPERIENCE

Data Science Intern, Imarticus **May 2024 - Present**

- Developed an advanced resume parsing tool using NLP for text extraction and LLMs for context-aware information extraction, enhancing candidate profile accuracy by 30%.
- Implemented grammar and spelling error detection with intelligent suggestions, improving the quality and professionalism of resumes by 40%.

Data Engineer, Pratham Software **July 2019 - Dec 2020**

- Designed and developed a CPQ (Configure Price Quote) tool tailored to specific customer requirements, enabling efficient and customized pricing strategies.
- Utilized Pyspark for automating data pipelines on Azure Data Factory (ADF), reducing delivery times by 30%

Trainee, Indian Space Research Organization **June 2018 - July 2018**

- Created a responsive website using MEAN stack for a Defense Laboratory within ISRO, demonstrating proficiency in full-stack development.

PROJECTS

SnapText: Image Chatbot

Developed a powerful and intuitive OCR application to detect and extract text from any kind of images.

- Utilizes advanced text recognition capabilities such as EasyOCR and PyTesseract to accurately extract text from images.
- Extracted text can be sent to a Chatbot to answer questions based on the provided text.

Technologies Used: Python, EasyOCR, OpenCV, Streamlit
GitHub Repository: SnapText

DataFlow Pro: Automating ML Workflows

Implemented a No-Code Python application to automate the end-to-end process of building, tuning, and evaluating machine learning models using JSON configuration files

- Automated ML pipeline for regression and classification tasks, improving efficiency by reducing 50% manual intervention.
- Implemented JSON-based configuration to reduce configuration errors and cut down analysis time by 40%.

Technologies Used: Python, Streamlit, Scikit-learn, JSON parsing
GitHub Repository: DataFlow Pro

ADDITIONAL INFORMATION

- **Certifications**:
  - Stanford | Deep Learning AI- Machine Learning Specialization
  - IBM - Machine Learning
- **Won** Imarticus Data Science Hackathon - 2024