Sonu Kumar Suman

Email : jsonusuman351@gmail.com | **Contact** : +919523235598 |Ranchi, Jharkhand |**Linkedin** : Link

PROFILE SUMMARY

Detail-oriented Senior Associate (Artificial Intelligence/Machine Learning) with 7+ months of experience in AI model training, data analysis, and NLP techniques (Natural Language Processing). Proficient in Python, pandas, NumPy, and TensorFlow/Keras for building scalable solutions. Expertise in data extraction, pattern recognition, and **cross-functional collaboration** to refine AI models for **Google Sunroof Solar Feasibility Prediction**, **Crop Disease Detection**, and **Skin Cancer Classification** (90% accuracy). Published researcher in **IoT-ML integration** for sustainable agriculture.

WORK EXPERIENCE

Senior Associate – AI/ML [Innodata India | SEP 2024 – FEB 2025]

1. **AI Model Training & Data Refinement**:
   - Trained and optimized generative AI models using MMD/PDF files and RWSW/LaTeX tools.
   - Applied NLP techniques (text cleaning, tokenization, NER) to extract structured data, improving model accuracy by **20%**.
   - Collaborated with **cross-functional teams** to automate **data preprocessing** workflows, reducing manual effort by **30%**.
2. **Google Sunroof Solar Feasibility Prediction**:
   - Analyzed aerial imagery using DataCompute tools to predict solar panel viability, reducing site visits by **40%**.
   - Integrated Python scripting for data validation and quality assurance.
3. **AI-Driven Crop Classification**:
   - Classified street-view images using Convolutional Neural Networks **(CNNs)**, achieving **93% accuracy**.
   - Resolved edge cases (occlusion, ambiguous land types) to enhance data labeling consistency.

MAJOR PROJECT :

Document Classification & Information Extraction System [*Personal Project |March* 2025 – *May* 2025][Github]

• Developed an end-to-end system to automatically classify scanned documents and extract key information using deep learning, OCR, and regex.

- Designed and implemented a **CNN** in **TensorFlow/Keras** for document type classification, achieving high accuracy.
- Built data pipelines for preprocessing and label encoding.
- Integrated Tesseract **OCR** and regex-based extraction modules.
- Deployed the solution as a **RESTful Flask API** for real-time use.
- Implemented SQLite-based logging for tracking predictions.
- Ensured modular code with documentation and error handling.

Key Technologies: Python, TensorFlow/Keras, OpenCV, pytesseract, Flask, SQLite, scikit-learn, Regex, pandas, NumPy

Precision Immunotherapy-Driven Early Detection System for Skin Cancer using Deep Learning [Kaggle|Jan '24 - May '24]

• Balanced a dataset of **5,000+ stereoscopic** images.
• Engineered a CNN system for lesion classification.
• Optimized training with TensorFlow/Keras and achieved >**97% precision, recall, and F1-score**.
• Deployed with containerization and API for real-time predictions.

Research Article| May 28 2024 IoT-enhanced machine learning for precise crop disease detection and sustainable agriculture (https://doi.org/10.1063/5.0215260)

Summary: Developed an innovative approach by integrating IoT sensors with Convolutional Neural Networks (CNNs) for early and precise detection of crop diseases, significantly advancing sustainable agriculture practices.

- **Integrated IoT sensors with CNNs for early detection, achieving >95% accuracy.**
- **Applied real-time analysis for global food security impact.**

SKILLS AND EXPERTISE

Programming: Python, Java, SQL | Data Structures & Algorithms

AI/ML: Machine Learning, Deep Learning, NLP, Keras, Data Preprocessing

Data Analysis and Visualization: Jupyter Notebook, Matplotlib, pandas, NumPy, Excel (Data Organization, Analysis)

Frameworks: TensorFlow, Keras, scikit-learn, XGBoost

Other: API Development (Flask), OCR (Tesseract), Regex, SQLite

EDUCATION

Degree	Institute	Score	Year
MCA	BIT Mesra, Ranchi, Jharkhand	CGPA: 8.5	2022–2024
BCA	T.N.B College, Bhagalpur, Bihar	74.08%	2017–2021
12th	Kendriya Vidyalaya, Kahalgaon	67.40%	2015–2017
10th	Kendriya Vidyalaya, Kahalgaon	CGPA: 7.2	2014–2015