Sawan Kumar

07987417089 ◇ sawanrawatjr10@gmail.com ◇ dehradun, uttarakhanad, India ◇ Open to Remote ◇ Open to Relocate ◇ LinkedIn ◇ GitHub

SUMMARY

Passionate about artificial intelligence, data science, and machine learning. Proficient in machine learning algorithms with hands-on experience in end-to-end pipelines and can conduct automating model life cycle including designing, implementing, and validating machine learning models, along with model optimization and monitoring post-production performance. Dedicated to continuous learning and development.

EXPERIENCE

Data Scientist Intern Jun '24 — Present
Kohli Media LLP dehradun, India (Remote)

Research Intern May '24 — Jul '24
Indian Institute Of Information Technology (IIIT) Dharwad, India (Remote)

- **Project: Monitoring Uttarakhand Weather for Cloud Burst Using CNN and GAF**
  - Developed a predictive model using Convolutional Neural Networks (CNN) and Gramian Angular Fields (GAF) to monitor weather conditions and predict cloudbursts.
  - Designed and implemented an API to fetch real-time weather data for various states and stations.
  - Created a FastAPI-based frontend for seamless interaction with the backend API and predictive model.
  - Conducted extensive data analysis and preprocessing to ensure accurate and reliable predictions.
  - Collaborated with a multidisciplinary team to enhance system functionality and user experience.

Infosys Summer Internship Jun '23 — Jul '23
Infosys Dehradun

- Achieved certification in Data Science, mastering data analysis, machine learning, and predictive modeling techniques. Completed an NLP certification, gaining expertise in analyzing, understanding, and extracting insights from human language data.
- Completed an NLP certification, gaining expertise in analyzing, understanding, and extracting insights from human language data.

PROJECTS

Transfer Learning for Medical Image Classification with EfficientNetB4 Link

- Developed a robust image classification model leveraging pretrained EfficientNetB4 architecture for binary classification of medical images. The project aimed to distinguish between benign and malignant cases with high accuracy, utilizing a custom dataset. Key highlights of the project include:
  - Collected real data from three different sources, Processed and augmented a dataset of 13,740 medical images (Train: 8,620, Validation: 3,060, Test: 3,060) into two classes (Benign: 7,956, Malignant: 5,784).
  - Added custom layers on top of the EfficientNetB4 base model, and fine-tuned the last 30 layers, enhancing the model's performance model architecture to suit binary classification tasks, enhancing model performance with domain-specific knowledge.
  - Achieved an AUC of 0.92, precision of 86%, and recall of 70% on the validation set at a 0.5 threshold. Adjusted the threshold to 0.32, improving recall to 80% while maintaining a precision of 80%.
  - Evaluation Results**:** On the test set, achieved precision of 80%, recall of 80%, and an F1 score of 0.80, with a sensitivity of 80.3% and specificity of 84.2% Generated precision-recall and ROC curves, demonstrating the model's trade-offs and performance across various thresholds.
  - Deployment Readiness**:** Ensured model robustness and reliability for real-world medical diagnostics, optimizing for high recall to minimize false negatives in critical healthcare applications.
- **Key Contributions**:
  - Achieved high classification accuracy, contributing to early and reliable diagnosis of medical conditions.
  - Enhanced model interpretability and transparency, ensuring trust and reliability in medical decision-making.

EnergyPro: Advanced Temporal Fusion Transformer for Accurate Power Demand Forecasting-(forthcoming)

Spam mail classifier - NLP and machine learning Link

- Developed a natural language processing model to classify emails as spam with 98% accuracy and 99.41% precision.
- Deployed the model on the Streamlit platform, showcasing practical application and deployment skills.

SKILLS

Programming Languages Python , R, Java, SQL
Frameworks Scikit-Learn, Tensorflow , Tensorflow extended (TfX), NumPy, Pandas, HuggingFace, LangChain, Excel
DevOps toolkit Docker, FastAPI, Git, MLflow, DVC, Google cloud platform (vertex AI, AutoML, BigQuery)

Concepts Data structures and algorithms , End to end machine learning operation for productions , Large Language Models, Statistics , NLP (RNN,LSTM,Transformer), Generative AI (GANs, VAEs, Diffusion Models), Explainable Artificial Intelligence, model optimization and monitoring post production performance

PUBLICATIONS

Jan '01

EDUCATION

Bachelor's degree in Computer science , Dehradun institute of technology (GPA: 7.39)
Sep '21 — May '25
Dehradun, India

Secondry education in , Army Public School (GPA: 8.5)
2018 — 2020
Gwalior, India