Sawan Kumar

Machine learning engineer

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.
  - Integrated alert system to notify relevant authorities and users of potential cloudburst conditions.
  - 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

Multi-Class Detection in Chest X-rays Using YOLO and VinDr-CXR Dataset., Kohli Media LLP Link dehradun, India

-
  - Worked with the VinDr-CXR dataset from a Kaggle competition, comprising over 15,000 annotated images.
  - Preprocessed DICOM files and converted them into suitable formats for training the YOLO model. Applied data augmentation techniques including blurring, grayscaling enhance model robustness.
  - Trained a custom YOLO model for detecting multiple classes of chest anomalies. Implemented anchor boxes and tuned model hyperparameters for optimal performance.
  - Used SGD optimizer with appropriate learning rates and weight decay parameters. Addressed memory issues by configuring PyTorch's CUDA memory management.
  - Evaluated model performance using precision, recall, F1-score, and mean Ensured model accuracy through cross-validation and testing on a validation set.
  - Utilized Python, PyTorch, YOLOv5 framework, Kaggle, Leveraged CUDA for efficient GPU utilization during model training.
- **Key Achievements**:
  - Successfully detected multiple classes of anomalies in chest X-rays with high accuracy.
  - Optimized data preprocessing and augmentation techniques to improve model robustness.
  - Demonstrated the ability to handle and process large-scale medical imaging datasets.
  - Applied advanced machine learning techniques to real-world medical data, showcasing practical experience and technical skills.

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
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

EDUCATION

Bachelor's degree in Computer science , Dehradun institute of technology (GPA: 8.2) Sep '21