Sajil Awale

Address: Huntsville, Alabama
Email :sajilawale@gmail.com
Mobile : +1-256-417-3690

Portfolio: www.sajilawale.com.np
Github: https://github.com/AwaleSajil
Linkedin: https://www.linkedin.com/in/sajilawale/

EDUCATION

- **University of Alabama in Huntsville** — Huntsville, Alabama ,USA
  Master's in Computer Science - Data Science (Concentration) — 2024 - 2026
  Current GPA: 4.0/4.0

- **Institute of Engineering, Pulchowk Campus, Tribhuvan University** — Lalitpur, Nepal
  B.E. Electronics & Communication; — 2016 - 2021
  Full Scholarship; Aggregate: 79.45%; Rank: 8th Position (Top 1% in University)

SKILLS

- **Languages**: Python, C++, C, C#, MATLAB, SQL
- **Machine Learning**: Pytorch, Transformers, Scikit-Learn, W&B, Spacy, Keras, OpenCV, Imbalanced-Learn, Hyperopt
- **Data Analysis Packages** : Pandas, Dask, Numpy, Scipy, Matplotlib, Seaborn, Plotly, NetworkX
- **Big Data Framework**: Pyspark, Hadoop
- **Frontend**: HTML, CSS, Bootstrap, JavaScript, Angular, jQuery
- **Backend**: Flask, Rest framework
- **Cloud Computing**: Amazon EC2, EMR Hadoop, EMR Serverless, Redshift, S3

EXPERIENCE

- **Graduate Research Assistant for LLM team**
  NASA-IMPACT @ UAH — *August 2024 - Present*
  - Science Keyword Recommender: Built an extreme multi-label classifier for NASA CMR, scaling from 430 to 3,240 science keywords. Used Focal Loss and custom stratified sampling to improve F1 to 0.55, enhancing metadata accuracy and dataset discoverability.
  - Pre-training Science Embedding Model (Indus-SDE): Pretrained a RoBERTa-based model on 520K NASA documents with extended 1024-token input and Weighted Keyword Based Dynamic Masking. Achieved 78.1% top-1 MLM accuracy, outperforming baselines on keyword tagging, astrophysics, and EJ tasks.
  - Downstream Task Unification Framework: Developed a modular multi-task fine-tuning pipeline using Hugging Face and W&B. Enabled plug-and-play config-based training/evaluation with automatic Excel reporting, streamlining model comparison and boosting team productivity.
  - Training Sentence Transformer (Indus-SDE-ST): Implemented DDP multi-GPU multi-stage training of a scientific sentence transformer using text/code pairs. Early results show superior performance on science-domain retrieval benchmarks, pushing forward scientific search and discovery tools.

- **Machine Learning Engineer**
  Cedar Gate Technologies — *July 2022 - July 2024*
  - Automated ETL field mapping by fine-tuning DistilBERT for multilabel classification to suggest source-to-destination field mappings and achieved 0.95 recall and 0.7 IoU. Initiated full ETL automation by fine-tuning Mistral-7B to autogenerate internal data transformation scripts.
  - Analyzed local model explainability tools (permutation SHAP, Deep Explainer, LIME), identifying FastSHAP as the optimal solution for a production diabetes model with extensive features based on speed and performance (87.2% Inclusion AUC).
  - Performed network analysis on healthcare providers to correlate patient-sharing patterns among physicians with medical costs for patients with chronic conditions like Chronic Heart Failure and Diabetes, revealing key cost drivers.
  - Optimized the segmentation of frequent ER visitors by systematically evaluating various scaling, feature extraction, and clustering methods. Utilized logistic regression coefficients for rapid cluster discrimination, ultimately identifying K-Means (6 clusters) with an auto-encoder as the most effective model, based on cluster metrics for overlap, quality, and cardinality.
  - Developed a LightGBM model to predict healthcare cost-risk (MARA scores), and the likelihood of it increasing or decreasing; achieving an R2 of 0.74 and MCC of 0.45, enabling proactive care management.

- Built a Gradient Boosted model to predict patient compliance with preventive care visits next year, achieving an MCC score over 0.75 to support targeted outreach initiatives.

Machine Learning Engineer
Docsumo   *March 2022 - June 2022*

- Benchmarked spaCy v2 vs. v3 Named Entity Recognition (NER) pipelines for information extraction from OCR-scanned documents based on their performance, speed, and size, providing key data for a strategic upgrade decision.

- Evaluated multiple document reading order detection techniques (e.g., DBSCAN, recursive XY-cut, layout reader, line-based block separation, docstrum) to enhance NER performance on complex layouts, measuring success with ROUGE-L and BLEU scores.

Associate Data Engineer
Deerwalk   *May 2021 - Feb 2022*

- Onboard new vendors (ETL processes on US healthcare data), ensure data integrity, analyse bugs which was triggered during data processing or client request and promptly resolve critical production issues

ACADEMIC AND PERSONAL PROJECTS

Funny Project
BigData Project   *2024*

- Engineered a two-stage NLP pipeline to classify 570,000+ jokes by humor, offensiveness, and sentiment, achieving a 0.86 weighted F1-score by fine-tuning a BERT model on a 55k-sample dataset labeled by local LLMs (Mistral, Gemma3).

Image Auto Alignment
Weekend Project   *2024*

- Built two solutions to auto-correct rotated images: Rule-based Flask API for documents (e.g., invoices) using line detection and text-weight heuristics. ML-based model with MobileNetV2 for general images; framed as a regression task and achieved 2.6° MAE on self-supervised Flickr dataset.

Real Time Visual Localisation and Mapping of Mobile Robot in Dynamic Environment
College Major Project   *2019 - 2020*

- A mobile robot capable of real-time Visual SLAM (Simultaneous Localization And Mapping) in a dynamic environment by reconstructing the entire 3D scene from 2D images captured by its camera. To address dynamic element, visual landmarks in dynamic areas are masked using ICNet, a semantic segmentation model fine-tuned to identify humans, the most prevalent dynamic objects.

Precision Livestock Farming — Improving Productivity of Broiler Chicken farm with technology
LOCUS 2019 Project   *2019*

- Project designed to monitor broiler chickens, utilizing YOLO for chicken detection and SORT (Simple Online Real-time Tracker) for mobility tracking. Eating behavior was estimated using a feeder microphone, while maintaining optimal environmental conditions, including temperature and humidity.

Vehicle Traffic Analysis and Management
College Minor Project   *2019*

- Traffic flow at various road junctions was assessed by vehicle counting with the help of YOLO and SORT from diverse originating sources. The Webster algorithm was used to determine the optimal timing for traffic signals.

Sajilomart
Everest Hackathon   *2019*

- Designed a prototype for effortless shopping: a seamless "grab and go" experience eliminating lines and checkouts, with automatic transaction handling.

Blind Eye — Assistive Technology for Blind People
Assistive Technology Hackathon   *2018*

- Designed a headset as a solution to enhance mobility for the visually impaired, aiding navigation and obstacle avoidance.

Exchange Program and Fellowship

- **Sakura Science Exchange Program**
  Japan Science and Technology Agency 16th - 23th Dec, 2019
  - Selected as one of the top 3 students for a program at Japan's National Institute of Technology, Kisarazu. We presented our poster, visited industries, and exchanged ideas and solutions with international peers.
  - Participated in sessions covering Japan's cutting-edge technologies, including Artificial Intelligence and the Internet of Things (IoT).

- **First Nepal Winter School in AI**
  Nepal Applied Mathematics and Informatics Institute for Research (NAAMII) 20th - 30th Dec, 2018
  - Learnt about probability and statistics, linear algebra, AI ethics, and Deep Learning through esteemed professors and guest speakers.
  - Finished hands-on lab assignments in computer vision and natural language processing (NLP).

Honors and Awards

- **Fonepay Student Ambassador**
  Fonepay   2020
  - Selected as one of the top 10 out of 100 competitive teams responsible for driving initiatives to promote and facilitate the growth of mobile payments.

- **Best Thematic Hardware Project**
  LOCUS   2019
  - We were honored to receive the award for 'Precision Livestock Farming' during the 16th edition of the National Technological Festival held by LOCUS, Pulchowk Campus

- **Institute of Engineering Scholarship for BE**
  Tribhuvan University, IOE   2017
  - Received full scholarship to study engineering in the most reputed engineering college of Nepal for securing 58th rank in competitive entrance examination given by more than ten thousand students.

Volunteering and Teaching experience

- **Training on ML Applications**
  Mentors Club, Cedargate   2023
  - Conducted comprehensive session for all Cedargate employees in Nepal covering a variety of ML algorithms, providing insights into our operational procedures and discussed both ongoing and completed projects that are currently contributing to our production

- **RoboPOP and Dronacharya Competitions**
  LOCUS   2020
  - Volunteered to develop 3D animations that would explain the regulations introduced for RoboPOP, an exciting new robotic balloon-popping event incorporated into LOCUS 2020. Additionally, I extended my commitment to creating animations for Dronacharya, a cherished and popular drone racing competition

- **Hardware Fellowship**
  LOCUS   2020
  - Instructed nearly 100 students, ranging from freshmen to sophomores, in the domains of Arduino programming and electronic hardware design
  - Mentored a team of junior-year students as they embarked on their project, creating a 2D CNC plotter designed for writing and drawing which was showcased at the 17th Technological Festival.

Online Certifications

- Deep Learning Specialization by DeepLearning.AI on Coursera.
- Applied Deep Learning Capstone Project by ibm on edx.
- Specialized Models: Time Series and Survival Analysis on Coursera
- Python Classes and Inheritance by University of Michigan on Coursera.
- Python (Basic) by Hackerrank