Diksha Aswal

+1 6073491503 | diksha.aswal94@gmail.com | diksha | Greater Boston Area, MA

Experienced Data Scientist proficient in Python, machine learning, and deep learning, with a track record of solving complex problems. Adept at effective communication and cross-functional collaboration, eager to contribute to innovative and challenging projects.

EDUCATION

The State University of New York at Binghamton, USA 08/2023 - 05/2025
Master of Science in Computer Science **CGPA: 3.94/4**
IIIT Allahabad, India 08/2018 - 06/2020
Master of Technology in Machine Learning and Information Systems **CGPA: 9.14/10**
Master's Thesis: Absolute Intersection Over Union: A complete Loss for Faster and Better bounding box regression.
UIET Kurukshetra, India 08/2013 - 06/2017
Bachelors of Technology in Computer Science and Engineering **CGPA: 8.0/10**

SKILLS

Languages : Python, SQL, C++, C, R
General Tools : Jupyter Notebook, Git, Rundeck, AWS (Cloud Development Kit, OpenSearch, S3, Glue, IAM, Lambda, SageMaker, RedShift), RStudio, DBVisualizer, MS Office Suite, MongoDB, Visual Studio Code
Data Visualization Tools : Google Charts, Looker, MS Excel(formulas, pivot tables), Tableau
Technical Skills : A/B Testing, Predictive Analysis (Decision trees, Clustering, Regression), Statistical Modeling and Analysis, Deep Learning (RNN, Attention, LSTM), Natural Language Processing, Graph API, Big Data, Computer Vision

EXPERIENCE

Amazon 08/2024 – 03/2025
Data Science Intern Boston, MA

- *Ticket Similarity Search:*Developed and deployed a ticket similarity model using OpenSearch and LLM-based embeddings to enhance ticket resolution efficiency. Built and deployed the model as an **AWS Lambda function** using **CDK**, integrating it with **API Gateway** for seamless accessibility. **Achieved an 89% average similarity score**, improving search accuracy and recommendation effectiveness.
- *Mitigation Tagging:* Developed an NLP based model to predict mitigation procedure tags for upcoming service tickets based on historical text data, **improved the performance by 20%**.
- Designed and implemented ETL pipelines to automate and streamline data workflows across modeling and analysis tasks.

Moonfrog Labs Pvt. Ltd. 09/2020 – 04/2023
Senior Data Scientist Bangalore, India

- Developed a revamped allocation method using gini coefficient measure, enhancing fairness in Teenpatti online game, yielding a **25% boost in competition equity**.
- Created Machine Learning based fraud detection algorithm, which **increased fraud detection by 133% per day**.
- Setup personalized promotions, leveraging player profile clustering, led to **12% revenue upsurge** in the initial phase.

Data Scientist

- Analysed data from **3M+ users**, suggest strategies to enhance KPIs post-feature launch.
- Created and optimized **ETL pipelines** to transform and load data into Redshift for reporting and analytics.
- Proficiently designed and streamlined various data **visualization dashboards, daily reports** to help identify critical KPIs and facilitate strategic planning.

Associate Data Analyst

- Performed **A/B testing** on gaming features like in-app promotions, **improved conversion rate by 10%**.
- Enhanced ML fraud detection on gaming platform, **reducing false positives by 64%**.

PUBLICATION

Aswal D., Shukla, P. Nandi, G.C. Designing effective power law-based loss function for faster and better bounding box regression. Machine Vision and Applications 32, 87 (2021). https://doi.org/10.1007/s00138-021-01206-5

PROJECT

Thesis Project 09/2019–06/2020

Developed a loss function for computationally efficient and accurate bounding box regression IIIT, Allahabad

- Studied bounding box regression loss functions for object detection using deep learning.
- Designed and integrated the novel AIoU loss function, outperforming l-norms, GIOU, DIOU, and CIOU.
- Evaluated performance via simulation experiments comparing computation time and error rates.
- Applied deep learning techniques and utilized state-of-the-art algorithms like YOLOv3 and SSD for object detection on the PASCAL VOC dataset.