

Diksha Aswal

+1 6073491503 | diksha.aswal94@gmail.com | diksha

Motivated Computer Science professional with strong skills in developing and deploying machine learning models, supported by cloud-native infrastructure on AWS. Proficient in Python, SQL, and C++, with experience building scalable applications, designing data pipelines, and solving complex technical problems.

EDUCATION

The State University of New York at Binghamton, USA

08/2023 - 08/2025

Master of Science in Computer Science

CGPA: 3.94/4

Research Project: Semantic Shift Detection in Congressional Hearings

Relevant Coursework: Natural Language Processing, Data Mining, Design and Analysis of Algorithms, Design Patterns, Artificial Intelligence, Database Systems

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

Relevant Coursework: Machine Learning, Deep Learning, Convex Optimization, Linear Algebra, Advanced Data Structures and Algorithms, Information Retrieval

UIET Kurukshetra, India

08/2013 - 06/2017

Bachelors of Technology in Computer Science and Engineering

CGPA: 8.0/10

Relevant Coursework: Data Structures, Programming Languages, Computer Architecture and Organization, Operating Systems, Computer Networks, Compiler Design

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 Retrieval System:** Built and productionized a cloud-native ticket similarity search model by integrating transformer-based embeddings (e.g., SBERT all-MiniLM-L6-v2, Amazon Titan) with AWS CDK, Lambda, OpenSearch, and API Gateway, enabling scalable retrieval with **89% similarity accuracy**.
- Mitigation Tag Classification Model:** Developed an NLP model to predict mitigation tags for service tickets with a large, imbalanced label space. Used TF-IDF features and a custom voting classifier (Logistic Regression, XGBoost, Random Forest, SVM), **improving performance by 20%**. Containerized with Docker and deployed via SageMaker, with S3 for data, ECR for images, and Lambda for inference.
- 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

- Improved competition equity by 25%** in the Teenpatti online game through fairness optimization using the Gini coefficient to reduce outcome inequality. Crafted the allocation model in Python using Pandas, NumPy, and visualized results with Matplotlib and Seaborn.

- Developed and deployed fraud detection models in a high-volume gaming environment using Isolation Forest for anomaly detection, **boosting daily fraud identification by 133%**. Engineered and managed the workflow using Python, Scikit-learn, and Amazon SageMaker.
- Setup personalized promotions, leveraging player profile clustering, and validated through A/B testing, led to **12% increase in in-app revenue**.

Data Scientist

- Performed exploratory data analysis (EDA) and funnel analysis on data from **3 Million+ users** to evaluate feature performance and uncover opportunities to improve key product KPIs.
- Designed and optimized ETL pipelines using Python, SQL, AWS Glue, and Rundeck to transform and load data from S3, RDS, and DynamoDB into Amazon Redshift, **improving data processing efficiency by 30%** and accelerating analytics workflows.
- **Created 20+ interactive and in-depth Looker dashboards** and streamlined daily reports to highlight key KPIs, enabling data-driven decision-making and strategic planning.

Associate Data Analyst

- Executed A/B testing on in-app promotions, **improving conversion rates by 10%**. Extracted experimental data using SQL, and conducted statistical testing and uplift analysis with Python (Pandas, SciPy, statsmodels).
- Improved ML-based fraud detection on the gaming platform by introducing a hybrid model using XGBoost, **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

Research Project

04/2025–Ongoing

Monte Carlo Simulations of 2D Circle Packing Efficiency

- Simulated 2D monolayers using physically realistic size distributions of circles to mimic granular or colloidal systems on a plane.
- Developed MC simulations at various length scales to analyze the effects of finite system size and boundary conditions on overall packing efficiency.
- Explored the configurational requirements to achieve Random Close Packing (RCP) and Random Loose Packing (RLP) critical benchmarks in disordered systems.

Research Project

04/2024–10/2024

Semantic Shift Detection in Congressional Hearings

SUNY, Binghamton

- Studied lexical semantic change (LSC) in Congressional hearing transcripts using self-supervised learning and global alignment techniques.
- Used contextual embeddings from BERT, RoBERTa, and TimeLMs to capture temporal shifts in word meaning.
- Measured semantic change via cosine similarity, centroid shift, and nearest-neighbor metrics.
- Tracked semantic evolution in policy language through visualizations using t-SNE, UMAP, and matplotlib; managed model development and evaluation with PyTorch and Weights Biases.

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.