MD ASJADULLAH

DATA SCIENCE/MACHINE LEARNING ENGINEER

+91 9852940253 | Kolkata, India | mdasjad895@gmail.com | LinkedIn: Md Asjad | Github |
Coding Ninja[master], Machine hack[155/10k], Kaggle, Leetcode[3*], hacker rank[5* SQL]

ACHIEVEMENTS

Analytics Olympiad 2023 **Rank:60/1000**
Developing machine learning models using credit history and payment behaviors data to predict customer loan default, addressing a critical issue financial institutions face in maintaining a healthy lending portfolio.

The Watermark **Rank:27/200**
Develop an AI-powered tool that can accurately classify images into two categories: with a watermark and without a watermark.

Data-Centric AI Competition 2023 **Rank:29/300**
Building multiclass-classification models for two data modalities (text and images) by utilizing the **clean lab** library.

Google Cloud Big Query LTV Prediction **Rank:27/900**
Used Big Query to build a predictive model for customers' Long Term Value (LTV) using anonymized real-world e-commerce data, involving data engineering, preprocessing, model building, and SQL.

WORK EXPERIENCE

Machine Learning Engineer Intern **10/2022 – 12/2022**
Technocolab Software Pvt.|Remote

- Pre-processed and cleaned a large and complex dataset to prepare it for EDA. Conducted Univariate, Bivariate, and Multivariate analyses to gain insights into data distributions, correlations, and trends.
- Collaborated with the team for Feature engineering. Developed and evaluated machine learning models for credit risk assessment, including Logistic Regression, Random Forest Classifier, Linear Regression, and Ridge Regression.
- Achieved an F1 score of 93.03 for the Random Forest Classifier model after feature selection.
- Conducted multi-target regression for EMI, ELA, and ROI prediction, achieving an R-squared (R2) score of 89 with Ridge Regression.
- Deployed machine learning pipelines for classification and regression tasks, facilitating model deployment and production use.
- Assisted in documenting model development and results, enhancing team knowledge and project reproducibility.

PROJECTS

A Secure Protocol for federated learning systems to protect against gradient attacks. | Federated Learning, Flower, PyTorch, CNN, Gradient Inversion, GPU **oct-2023**
Team Project, IIITDM KURNOOL ***Github***

Collaboratively we have to design our architecture by inspiration from research previously done that is better in gradient inversion attacks in federated learning [**Multiclass classification**]. I worked on the survey of different attacks against the gradient and implemented our proposed architecture and Gradient inversion attack.

- Defined and implemented a **custom semantic loss function** for autoencoder training, improving robustness against gradient inversion attacks.
- Successfully implemented **semantic loss** and **gradient inversion attack** on MNIST/CIFAR-10, achieving state-of-the-art results with metrics like **PSNR, MSE, and SSIM** in reconstructing images from gradients. of clients.
- Leveraged trained autoencoders with semantic loss to train/test ImageNet classifier, achieving **97%** accuracy on original data vs. **83%** on autoencoder output.
- Collaborated with a team of 3 members and 2 professors on a novel approach to improve federated learning.

VidSumAI|Javascript, HTML/CSS, Python, Flask, BART, Whisper, Transformers, GPU **Github**

- Developed a Flask-based web app for **video summarization, transcription, audio summaries, and translation**. Fine-tuned the **BART(facebook/bart-base**) model for video summarization on **lighteval/summarization** data.
- Used Modular coding for steps like data ingestion, data validation, data preprocessing, etc. Evaluated algorithm performance using **BLEU and ROUGE scores**.

- Future plans include keyword extraction, video segmentation, and playlist generation for further improvements.

MarketMoodMeter| Python, Streamlit, Data analysis, Time Series, NLP, Finance **Github**

- Created a multi-page Streamlit App for analyzing market news headlines alongside stock prices. Utilized data scraping and fetching from 2 distinct sources for news articles and company details, while simultaneously pulling real-time stock data from yfinance based on a user-provided ticker.
- Performed ETL (Extract, Transform, Load) operations as part of the data processing. Performed in-depth trend analysis on hourly and daily data, extracting vital statistics for significant days.
- Identified influential news articles and created a stock price forecasting model with sentiment scores, conducted thorough analysis, including correlations, and deployed it on Streamlit Cloud for user-friendly access.

ContentCompanion| Python, Streamlit, Django, CNN, Image processing, computer vision, GPU **Github**

• Developed a content-based search engine powered by a Siamese neural network and employed few-shot learning techniques to enable efficient image similarity recognition.
• In Model 1, a Siamese neural network architecture was crafted, consisting of 3 convolutional layers, 3 max-pooling layers, and 3 dropout layers, and 1 global average-pooling layer Likewise, Model 2 featured a parallel setup with an identical configuration.
• Achieved 0.05 validation loss in 50 epochs, closely monitored on GPU, utilizing 560 training and 140 testing samples with Adam optimizer and contrastive loss.

EDUCATION

Bachelor's Degree in Artificial Intelligence and Data Science **2020 – 2024**
Indian institute of information technology design and manufacturing Kurnool, India [IIITDM KURNOOL]
High School Bela Shanti Kutir, Bihar **2019**
Senior Secondary Certificate, Bihar School Education Board **(**B.S.E.B**)** **80%**
IIIT JEE Main 2020 **97%ile**

PROFESSIONAL DEVELOPMENT

Google Data Analytics Professional Certificate [Coursera] ||Natural Language Processing [NPTEL]
Machine Learning [Coursera] ||Reinforcement Learning [NPTEL]
Deep Learning Specialization [Coursera]
IBM Data Engineering Professional Certificate [Coursera]

TECHNICAL SKILLS

Languages: C++, Python, R, HTML/CSS, JavaScript
Libraries: C++ STL, TensorFlow, Keras, PyTorch, scikit-learn, OpenCV, NumPy, Pandas, Nltk, Flask, Django, Cudnn
Tools: VScode, Git, Github, PyCharm, Spyder, JupyterNotebook
Cloud/Databases: Relational Database (MySQL), MongoDB
Relevant Coursework: Data Structures and algorithms, Operating Systems, Object Oriented Programming, Database Management Systems, Software Engineering, Machine Learning, Computer vision, Natural Language Processing, Data Analysis and Visualization, Probability and Statistics, Reinforcement Learning, Deep Learning.
Soft Skills: Teamwork, Time Management, Leadership, Problem Solving, Self-learning, Presentation, Adaptability