Dhruvraj Singh Rathore

dhruvraj_16@tamu.edu| LinkedIn | 7372061179

EDUCATION

Texas A&M University, College Station, TX **August 2024 – December 2025**
Master of Science in Data Science

SRM Institute of Science and Technology, Chennai, India **July 2018 – May 2022**
Bachelor of Technology in Computer Science Engineering

SKILLS

Programming:	Python, R Programming, Pyspark, Pandas, Numpy, Pytorch, Sklearn, Tensorflow
Visualization:	Power BI, Matplotlib, Seaborn
DataBase:	SQL, Postgres, NoSQL, MongoDB, Data Warehouse, Redis
Cloud computing:	AWS Suite (EMR, IAM, S3, EC2, Glue, Lambda), Apache Spark (PySpark), DevOps

PROJECTS

Cotton Field Detector

- Developed an automated method to identify and map cotton crop areas in the United States using satellite imagery from Google Earth Engine, where each pixel represented one of 254 different crops.
- Implemented image processing techniques to generate multiple smaller images from the large satellite image and created binary masks highlighting cotton areas while masking other crops.
- Employed UNET algorithms in **PyTorch** to perform **semantic segmentation** on the processed images, generating predicted mask images from input images.
- Evaluated the model's performance by comparing predicted masks with true masks using Dice coefficient and IoU (Intersection over Union) metrics.
- Calculated the area covered by cotton crops by analyzing pixel data and converting it to acreage, demonstrating the application of deep learning techniques in agricultural remote sensing.

Metastasis Detection in Histopathological Images Using Convolutional Neural Networks (CNNs)

- Utilized the PCam dataset, which consists of 327,680 color images (96x96 pixels) extracted from lymph node histopathologic scans.
- Implemented a custom Convolutional Neural Network (CNN) specifically designed for this binary classification task, leveraging the spatial hierarchies in histopathological images.
- Employed data augmentation to enhance model generalization and prevent overfitting. Utilized batch processing for efficient training and memory management. This Mainly included operations like random horizontal/Vertical rotations
- Applied the Adam optimizer to adjust the model's parameters during training, known for its effectiveness in training CNNs.
- Achieved a notable F1 score of 0.8768, demonstrating the CNN's strong performance in balancing precision and recall for metastasis detection in histopathological images.

Metro Interstate Traffic Volume

- Developed a machine learning model to predict traffic congestion using regression techniques, focusing on multiple factors such as time of day, weather conditions, and historical traffic patterns.
- Employed data preprocessing techniques such as feature engineering, handling missing values, and normalizing data to improve model performance and capture complex relationships between variables.
- Implemented and compared multiple regression algorithms, including Linear Regression, Multiple Linear Regression Forest, Polynomial Linear Regression with different regularization techniques like Lasso and Ridge to determine the most effective approach for traffic congestion prediction.
- Evaluated model performance using metrics like Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-squared (R2) score. Polynomial model performing the best among the basic models without any additional t uning.