Shreya Patil

Mobile +1-443-453-8281 | shreya.patil.u@gmail.com |Baltimore, MD

https://www.linkedin.com/in/shreya-patil-470219105/ http://www.shreyapatil-datascientist.com/

PROFESSIONAL SUMMARY

- Data Scientist with 3+ years of experience in **Machine Learning, Deep Learning and High-Performance Computing** technologies with enhanced skills of building prototypes and designing algorithms.
- Machine Learning Researcher with a demonstrated history of working on Weather Forecasting using Neural Net Architecture Search Algorithm.
- Experience with tools **Tableau, Jupyter, GitHub.**
- Proficient in **Python (Pandas, NumPy, NLTK, Matplotlib, Scikit-learn, PyTorch, Tensorflow), PySpark, SQL, Linux, CUDA.**
- Professional experience with Data warehouse, Data Pipelines, Exploratory Data Analysis.
- Successful background of managing Databases and Cloud Platforms by leveraging infrastructure and operational knowledge for four clients in past work experience.

PROFESSIONAL EXPERIENCE

Research Assistant, UMBC, MD (PI: Dr. Milton Halem) Aug 2022- Present

- Developing a Wildfire Digital Twin (WDT) model, model for the simulation of fire, its spread and predicting the smoke impacts on air quality.
- Evaluated Global Tree Cover Loss to predict the impact of mega wildfires and deforestation on the irreversible climate change in tropical forests.
- Acquired 25 TB fire data (Grib files) from the HRRR instrument on the NOAA satellite to train the ResNet model.

Data Scientist, Redapt Inc May 2022- Aug 2022

- Build a predictive model for Make-a-Wish organization using Decision Tree on Azure ML Studio for effective resource management.
- Collaborated closely with clients to ascertain data architecture and analysis needs to design custom analytics reports.
- Analyzed data from Azure Blob Storage to create reports using PowerBI tool.
- Researched and prepared a presentation on Quantum Computing for FinovateFall conference.

Research Assistant, UMBC, MD (PI: Dr. Milton Halem) Aug 2021- Apr 2022

- Developed a Deep Neural Network architecture using python for time series microphysics parameters, to reduce the computational time by replacing components in NASA Unified-Weather Research Forecasting model (NU-WRF).
- Trained and tested Auto-Kera's Neural Net Architecture Search (NAS) algorithm on WRF-CHEM models microphysics output over US nested domains at a coarse and high-resolution data with RMSE of 0.015.
- Analyzed 3 TB data stored in Hadoop distributed file system to retrieve essential parameters of microphysics to predict the precipitation with 40% speedup.

Senior Analyst, Capgemini Technology Service LTD, INDIA Mar 2018- Nov 2019

- Maintained and administered databases and virtual environments in VMware and Microsoft Azure platform for 10 clients.
- Delivered in-depth training, imparting knowledge of best practices to the 10+ new recruits as Account Lead.

- Organized system infrastructure documentation and operating procedures, strengthening overall team performance.

EDUCATION.

Master of Professional Science - Data Science **May 2023**
University of Maryland, Baltimore County (UMBC), Baltimore, MD

Master of Technology - Computer Science and Engineering **May 2019**
Shivaji University, MH, INDIA

Bachelor of Engineering – Computer Science and Engineering **May 2017**
Shivaji University, MH, INDIA

PUBLICATION AND CONFERENCES

- Accelerating the WRF-CHEM Model Using a Machine Learning Emulation *Poster*, 2022 American Meteorological Society Conference.
- Patil S U (2019), Optimal Number of Cluster Identification using Robust K-means Algorithm. International Research Journal of Engineering and Technology.
- Patil S U (2018), A Review of Clustering and Clustering Quality Measurement. International Research Computer Engineering in Research Trends.
- Technical Content Creation *Medium*

PROJECTS

Social Bias Indicator **May 2022**
Trained Open-GPT2 (Generative Pre-trained Transformer) and Open-GPT3 models to predict the social bias in the online social media posts. This model classifies the biased implications into categories like - Offensiveness, Stereotype and Biases which can be further used to debias the datasets before training the NLP models.

Histopathological Cancer Detection **May 2022**
Designed and implemented VGG, AlexNet, and LeNet CNN models for histopathological cancer detection. Preprocessed the dataset, performed data augmentation techniques, and trained the models with the same hyperparameters. VGG achieved the highest accuracy and F1-score, followed closely by AlexNet and LeNet. Visualized the learned features of the models using activation maximization.

Prediction of Drug Binding Affinity of Protein Using Spark ML **May 2022**
Extracted biological data for a single protein target. Used PySpark classification models to predict binding between sample drug compounds and target protein. Compared multiple models to find optimum solution for prediction.

Understanding the X-ray Diffraction Data of Protein Structures to Predict Resolution **Mar 2022**
Filtered X-ray diffraction data for available protein structures from RCSB Protein Data Bank. Compared Logistic regression with regularization and decision tree models with hyperparameter tuning using grid search cross validation method. Predicted the resolution of protein structure with 85% accuracy.

Optimal Number of Cluster Identification using Robust K-means Algorithm. **May 2019**
Developed a Robust K-means Algorithm to identify optimal numbers of clusters in protein sequences by removing noise clusters. Measured goodness of clusters using Silhouette Coefficient.