Pragya Jatav

pragyajatav11@gmail.com | (+91)7522092221 | linkedin.com/in/pragya42 | pragya42.github.io

Skills

Programming Languages

Python • R• SQL• Spark

Technologies and Frameworks

• Pyspark • StreamLit • Advanced Excel • Google Cloud •Tensorflow • GIT• Tableau • MLFlow

Libraries

OpenCV• pandas• numpy• sklearn• nltk

Education

Indian Institute of Technology, Kanpur

2016 – 2020

Bachelors of Technology in Materials Science and Engineering, Minor in English Literature

Courses and Certifications

Probability and Statistics
Deep Learning by Open AI
Natural language processing
Introduction to data engineering
GCP Associate Cloud Engineer
Data Structures & Algorithms

Capstone Projects

• Used India GIS Data to visualize distribution of mineral ores in India.
•Created a Tableau Dashboard depicting customer segmentation and campaign metrics for a fashion retailer
•Developed an encoder(CNN) and decoder(RNN) model with attention to generate captions on images.
•Identified potential churn customers, reduced data imbalance, conducted correlation analysis and outlier removal and developed a Logistic Regression Classifier

Interests and Volunteering

Taught Mathematics to 15 underprivileged children at Prayas, student run initiative at IIT Kanpur

AI/ML mentor at Masai school and topmate.io.

Reading, Blogging, Film Criticism, Graphic Design, History

Work Experience

Blend 360

Data Scientist | Jul'22 – Present

- Created and analyzed referral networks for a fin-tech firm, developing a K-means clustering algorithm using PySpark and SQL to distinguish risky referral networks.
- Worked on a risk score framework to assess customer risk across the user journey. Analyzed transaction and customer metadata, developing an XGBoost model to identify risky customers. Deployed the model using MLFlow and Databricks workflows.Developed an automated retraining framework to keep the model relevant with evolving data trends.
- Explored various methods for developing a recommendation engine tailored for a home improvement company. Used python and SQL to implement three probabilistic and one deep learning solutions.
- Used RFM, behavioral and demographic data to create customer segments for a fashion retailer. Used SQL to create a datamart and DBSCAN clustering for segmentation.
- Created a demand elasticity model using ridge regression. Optimsed discounts by the predicted demand using genetic algorithm for a retailer.

ICICI Lombard GIC

Manager, Business Intelligence | Sep'20 – Nov'21

- Worked in the Business Intelligence team to develop and deploy solutions using data analysis, machine learning and deep learning techniques coupled with Natural Language Processing and Computer Vision.
- Analysed website data from Google Analytics for Lead prioritization
- Evaluated (using CSI and PSI) , retrained and monitored the previous classification model,deployed the same using Azure data factory.
- Developed an end to end solution for profiling and risk assessment of intermediaries
- Preprocessed images and developed a transfer learning model for image classification
- Streamlined the process to convert audio file types using python.
- Received an accolade from Senior data scientists for analysing multiple speech to text transcription services in a short duration.

Projects

Predicting water level of a Lake | Time Series Forecasting

- Forecasting the water level of a lake, using data from the Acea Smart Water Analytics challenge .Pre-processed dataset by handling missing values and resampling
- Used ADF test to check for stationarity, performed feature engineering and autocorrelation analysis,generated predictions by using ARIMA model with MSE 2.5

Autocorrection Feature | Natural Language Processing

- Developed my corpus of words from previous social media posts, blogs and chats.
- Utilized textdistance for similarity computation and Jaccard distance for autocorrect suggestions

Book Recommendation System | Recommender Systems

- Developed a book recommendation systems utilizing collaborative filtering techniques including Memory-based and Model-based approaches.
- Conducted EDA on the Book-Crossing dataset. Developed models for both approaches utilizing KNNWithMeans (item-based) and SVD (matrix factorization).

Academic Projects

Prepration of Molybdenum disulphide Quantum Dots

Prof. Krishanu Biswas, MSE IITK | May'19 - Jul'19

- Prepared MoS2 Quantum Dots via liquid phase exfoliation. Utilized cryomilling for nanoparticle synthesis and ultrasonication for quantum dot formation.
- Verified via multiple spectroscopic techniques. Successfully produced MoS2 Quantum Dots ranging from 18 nm to 50 nm.