Pragya Jatav

 pragya42 | pragya42 | pragyajatav11@gmail.com | +91-7522092221

WORK EXPERIENCE

Blend 360-Data Scientist   Jul 2022 - present

- **Discount optimization (Python, Genetic algorithm)**
  - Created a discount optimization framework to create discount strategies for an e-commerce firm.
  - Created a demand forecasting model using ridge regression to predict demand and Genetic algorithm to calculate optimum discounts to generate maximum revenue
  - Increased the average weekly revenue by 7% and increased stock depletion rate by 5%
- **Personalized email generator (Python, GPT 3.5)**
  - Created a tool that develops hyper-personalized marketing emails for users
  - Used clustering algorithms to develop hyper-personalized marketing emails to users
  - Generated product recommendations and created personalized emails by Retrieval-Augmented Generation using GPT 3.5 turbo
  - Created a discount optimization framework to create discount strategies for an e-commerce firm.
  - Recommended products with 95% accuracy and increased email response by 10%
- **Customer Risk Framework(Pyspark, Databricks, ML Flow, SQL)**
  - Developed a risk score framework consisting of a suite of models and business rules to analyse customers.
  - Used SQL and databricks to prepare and analyse customer, transaction and demographic data.
  - Developed business rules, unsupervised and supervised models using Pyspark to calculate customer risk score at every step of user journey. Used MLFlow to deploy the same
  - Increased risk evaluation efficiency by 23%, fulfilled compliance and anti-money laundering requirements, reduced customer risk evaluation time from 10 days to 3 hours.
- **Recommendation Engine (SQL, Tenserflow)**
  - Developed a repeat purchase recommender system for a home improvement company
  - Implemented session-based recommender system utilizing GRU session encoding, attention mechanisms, and mode prediction for enhanced item recommendations. Deployed the model using GCP
  - The deep learning-based recommender system resulted in 15% boost in conversion rates.
- **Customer Segmentation Analysis (DBSCAN Clustering, SQL, Tableau)**
  - Developed a customer segmentation model and dashboard to analyse and acquire customers.
  - Used RFM, behavioral and demographic data to generate insights. Created customer segments using DBSCAN clustering. Created a Tableau dashboard to showcase insights.
  - Created a lookalike model using KNN clustering to find potential customers.

ICICI Lombard Geneal Insuance Company- Data Scientist   Jul 2020 - Nov 2021

- **Google Analytics Lead prioritization (XGBoost, Azure Data Factory)**
  - Created datamart from Google analytics data, conducted EDA to generate insights.
  - Evaluated previous model using CSI. Developed an XGBoost classification model for potential customers
  - Increased average monthly lead generation by 9 percent
- **Insurance Claim Report Analysis (Tenserflow, Tesseract)**
  - Develop a tool that checks all resources have been filed for health insurance claims.
  - Developed an object detection model using transfer learning to check if all forms have been filed.
  - Developed an OCR model using Tesseract OCR to extract and cross-check details from the forms

SKILLS

Languages	Python, R, SQL, Spark
Frameworks	Pyspark, StreamLit, Advanced Excel, Google Cloud, Tensorflow, GIT, Tableau, MLFlow
Libraries	OpenCV, pandas, numpy, sklearn, nltk

COURSES AND CERTIFICATIONS

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

EDUCATION

2016 - 2020   B.Tech (Materials Science and Engineering) at **Indian Institute of Technology, Kanpur**

ACADEMIC PROJECTS

Prepration of Molybdenum disulphide Quantum Dots
Prof. Krishanu Biswas, MSE IITK

- 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.

PROJECTS

- **Predicting water level of a Lake (Time Series Forecasting)**
  - Forecasting the waterlevel of a lake, using data from the Acea Smart Water Analytics challenge .Preprocessed 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).

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 outlierremoval and developed a Logistic Regression Classifier