Pragya Jatav

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

WORK EXPERIENCE

Blend 360-Data Scientist   Jul 2022 - present

- **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, Tensorflow)**
  - Led a team of 3 to develop a repeat purchase recommender system for an E-commerce firm
  - 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.
- **Personalized email generator (Python, GPT 3.5)**
  - Created a tool that develops hyper-personalized marketing emails for users
  - Used clustering algorithms to create customer segments, Generated product recommendations and created personalized emails by Retrieval-Augmented Generation using GPT 3.5 turbo
  - Recommended products with 95% accuracy and increased email response by 10%
- **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 (Tensorflow, 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**