LinkedIn
jainpalak3286@gmail.com

Palak Jain

Website

Salt Lake City, UT

Data Scientist with over 4 years of experience in analytics. Proficient in developing machine learning and neural network models to extract actionable insights from complex datasets. Expertise in NLP, computer vision modeling, and scalable model deployment.

Education

Master of Science, Applied Data Science, Indiana University Indianapolis, **(GPA: 3.8/4.0)** **Jan 2021 - Dec 2022**

Bachelor of Engineering, Electronics and Telecommunication, Devi Ahilya University, India, **(GPA: 7.2/10.0)** **July 2014 - May 2018**

Experience

Data Scientist **Feb 2023 - Current**

Verisk- Claims Insight and Analytics **Lehi, UT**

- Spearheaded the development and implementation of advanced machine learning models for fraud detection in insurance claims achieving 94% precision and reducing false positive rates to 2%.
- Collaborated with cross-functional teams including software engineers and domain experts to deploy the model, integrate real-time data streams and enhance fraud detection efficiency by 40%, for timely fraud prevention.

Data Science Excellence Program Intern **June 2022 - August 2022**

Verisk- Claims Insight and Analytics **Jersey City, NJ**

- Implemented multiple CV models to detect text and face in scene images to protect Personal identifiable information for claims data.
- Proposed an OpenVino tookit based computationally efficient CV model with high recall of 0.85 and speed of ~20 images per second.
- Business Impact Project: Saved cost of resurveying 20K+ commercially insured underwriting by extracting and combining information from Verisk and its acquired companies.

Research Assistant || National Science Foundation(NSF) **Jan 2021 - Dec 2022**

Indiana University Purdue University Indianapolis **Indianapolis, IN**

- Extracted Causal relationships from 1 million+ sentences using semantic and syntax cues and captured the strength of the relationship.
- Created RNN model using BiLSTM and PyTorch to determine contextual information in both forward and backward directions with Receiver Operator Characteristics (ROC) of 0.98.
- Employed pre-trained language models for cause-effect extraction, fine-tuned with CauseNet data which improved F-score by 8%.
- Discovered fact localization and edited BERT's parameter to update incorrectly predicted or obsolete facts.

Data Engineer **Jan 2019 - Nov 2020**

Infosys Limited **Hyderabad, India**

- Catalogued financing data for e-contract utilization from Online Transaction Processing (OLTP) servers and flat files using Informatica.
- Designed and tested 350+ Extract Transform Load (ETL) Mappings, Workflows using Informatica Powercenter with 150+ tables.
- Optimized SQL queries for unit testing, enhanced performance of ETL process by 200% with partitioned tables and parallel processing.

Technical Skills

Languages	Python (Pandas, NumPy, Scikit-learn, TensorFlow, PyTorch), R, SQL, HTML
Database and Cloud	RDBMS, MySQL, SQL Server, Informatica, Microsoft Access, AWS S3
Analytics Tools	Power BI, Tableau, Matplotlib, Microsoft Excel
Statistical Skills	Statistical Modeling, Hypothesis Testing, Predictive Modeling, Exploratory Data Analysis, Data Visualization, ML, DL, Data Mining, Neural Networks, Word Embedding, Dimension Reduction, Parameter Optimization, Natural Language Processing (BERT, SpaCy, NLTK), Computer Vision(OpenCV, OpenVINO)
Model Deployment	Flask, Docker, Heroku

Projects

Pre-Owned Car Market | Data Visualization | Prediction Model | PowerBI | Flask | Heroku

- Created interactive PowerBI dashboard to visualize and investigate car price variation with 10+ features of the car.
- Deployed prediction model on Heroku cloud platform for online estimation of pre-owned car prices.

Diabetes Onset Prediction | NLP | Pytorch | Artificial Neural Networks(ANN) | GPU

- Determine the patient's diabetic condition based on rapidly diagnosable measures including Blood Pressure, Glucose level, and BMI.
- Developed an ANN classification model using PIMA Indian Diabetes Dataset and PyTorch framework resulting in accuracy of 80.5%.

All Other Projects

Publications

1. VanSchaik, J. *et al.* Using transfer learning-based causality extraction to mine latent factors for Sjögren's syndrome from biomedical literature. *Heliyon* (2023).