

Palak Jain

LinkedIn

jainpalak3286@gmail.com

Website 

Salt Lake City, UT

Data Scientist with 5+ years of experience driving business impact through advanced analytics and neural networks. Expertise in NLP, computer vision, and scalable model deployment, delivering actionable insights that power data-driven decisions.

EXPERIENCE

Data Scientist - Claims Fraud Detection

Feb 2023 - Current

Verisk Analytics

Lehi, UT

- Built a real vs. AI-generated image detection model, curating refined datasets from multiple image generation tools; achieved 80% precision with <2% false positives, and delivered a Gradio-based demo app containerized via Docker for production deployment.
- Designed and deployed advanced image forensics model to detect pixel-level edits in insurance claim images, achieving 94% precision with <1% false positives, and implemented heatmap-based localization to highlight manipulated regions.
- Led end-to-end development of a 32-class image classification model using CLIP embeddings to filter irrelevant categories before downstream fraud detection checks; successfully integrated into the existing fraud detection pipeline.

Data Science Intern

June 2022 - August 2022

Verisk Analytics

Jersey City, NJ

- Implemented multiple Computer vision models to detect text and face in scene images, safeguarding personally identifiable information (PII) in insurance claims data.
- Designed a computationally efficient CV model using OpenVINO Toolkit, achieving 0.85 recall and processing speed of ~20 images/second, enabling faster and cost-effective large-scale screening.
- Business Impact Project: Reduced the need for resurveying 20K+ commercial underwriting cases by building an automated data integration pipeline that extracted and merged information from Verisk and its acquired companies, resulting in significant cost savings.

Research Assistant || National Science Foundation(NSF)

Jan 2021 - Dec 2022

Indiana University Purdue University Indianapolis

Indianapolis, IN

- Extracted causal relationships from 1M+ sentences using semantic and syntactic signals and quantified the strength of each relationship.
- Developed a BiLSTM RNN model using PyTorch to capture bidirectional context, achieving a ROC score of 0.98.
- Fine-tuned pre-trained language models on CauseNet data for cause-effect extraction, improving F-score by 8% over baseline.
- Enhanced fact localization by modifying BERT parameters to correct inaccurate or outdated predictions.

Data Engineer

Jan 2019 - Nov 2020

Infosys Limited

Hyderabad, India

- Catalogued financing data for e-contract utilization from OLTP servers and flat files using Informatica.
- Designed, developed, and tested 350+ ETL mappings and workflows in Informatica PowerCenter, managing data across 150+ tables.
- Optimized SQL queries for unit testing and improved ETL pipeline performance by 200% through partitioning and parallel processing techniques.

TECHNICAL SKILLS

Languages	Python, SQL, HTML
Database and Cloud	RDBMS (MySQL, SQL Server), ETL (Informatica PowerCenter), Cloud (AWS S3, AWS EC2)
Analytics Tools	Power BI, Microsoft Excel (Advanced), Matplotlib, Seaborn
Statistical Skills	Statistical Modeling, Hypothesis Testing, Predictive Modeling, Exploratory Data Analysis, Data Mining, Parameter Optimization
Machine Learning and Deep Learning	Supervised & Unsupervised Learning, Neural Networks, CNNs, RNNs, BiLSTM, Word Embeddings (Word2Vec, GloVe), Dimension Reduction
Natural Language Processing (NLP)	BERT, SpaCy, NLTK, Cause-Effect Extraction, Causal Inference
Computer Vision	OpenCV, OpenVINO Toolkit, Image Classification, Object Detection, Heatmap Localization
Model Deployment	Flask, Docker, Gradio, Cloud Deployment (AWS EC2)

EDUCATION

Master of Science, Applied Data Science, Indiana University Indianapolis

Jan 2021 - Dec 2022

Bachelor of Engineering, Electronics and Telecommunication, Devi Ahilya University, India

July 2014 - May 2018

PERSONAL 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).

