LinkedIn
paljain@iu.edu

Palak Jain

Website
Indianapolis, IN

Data Science enthusiast with 2+ years of experience in Analytics. Proficient in creating machine learning and neural network models using Python. Hands-on experience in database design, SQL, and visualization tool PowerBI for identifying patterns and extracting valuable insights.

Education

Master of Science, Applied Data Science, Indiana University Purdue University Indianapolis, **(GPA: 3.8/4.0)** **January 2021 - Ongoing**
Bachelor of Engineering, Electronics and Telecommunication, Devi Ahilya University, India, **(GPA: 7.2/10.0)** **July 2014 - May 2018**

Experience

Data Science Intern **June 2022 - August 2022**
Verisk Analytics **Jersey City, NJ**
- Implemented several CV models to detect text and face in scene images to protect Personal identifiable information for claims data.
- Proposed an accurate, fast and easily deployable model with minimum resources, achieved recall of 0.85, speed of 15-20 frames per second and full compatibility with CPU.
- Analyzed commercial property insurance underwriting survey data of past 2-3 decades, updated it using data of different companies acquired by Verisk, eventually saved cost of resurveying for more than 20K properties.

Research Assistant || National Science Foundation(NSF) **January 2021 - Ongoing**
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.
- Constructed a language model using named entity recognition with Spacy, word-embeddings with Gensim Word2Vec, and dimension reduction techniques PCA and TSNE to analyze vector representation of words.
- 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.
- Improved F-score of model by replacing existing rules-based approach (82%) from deep learning approach BiLSTM (94%).

Data Engineer **January 2019 - November 2020**
Infosys Limited **Hyderabad, India**
- Catalogued the extracted data from Online Transaction Processing (OLTP) servers and 10+ flat files using Informatica to assess e-contract utilization for financing.
- Designed, developed, and tested 350+ ETL Mappings, Workflows, Worklets using Informatica Powercenter with 150+ tables.
- Optimized SQL queries for unit testing and enhanced performance of ETL process by 200% using partitioned tables and parallel processing.

Technical Skills

Languages	Python, R, Core Java, HTML
Database Management	RDBMS, MySQL, SQL Server, Big Data, Hadoop, Informatica, Microsoft Access
Analytics Tools	Power BI, Tableau, Microsoft Excel, Microsoft PowerPoint
Statistical Skills	Hypothesis Testing, Statistical Modeling, Predictive Modeling, Exploratory Data Analysis, Data Visualization, ML, DL, Data Mining, Neural Networks, Natural Language Processing, Word Embedding, Dimension Reduction, Parameter Optimization, Computer Vision. (Libraries and Frameworks: Numpy, Pandas, Scikit-Learn, NLTK, Gensim, Spacy, OpenCV, Pytorch, Tensorflow)

Projects

Pre-Owned Car Market | Data Visualization | Prediction Model | PowerBI | Flask | Heroku
- Created PowerBI interactive dashboard to understand variation of price with more than 10 different features of the car.
- Modeled linear regression to predict car price and deployed it on the cloud platform.

Diabetes Onset Prediction | NLP | Pytorch | Artificial Neural Networks(ANN) | GPU
- Developed classification model to determine the diabetic condition of patients based on rapidly diagnosable measurements including Blood Pressure level, Glucose level, and Body Mass Index (BMI).
- Built an ANN model with PIMA Indian Diabetes Database using PyTorch which resulted in an accuracy of 80.5%.

Tweet Sentiment Extraction | NLP | Keras | Word Embeddings | Anaconda | Recurrent Neural Network(RNN)
- Developed RNN model with Functional API to analyze polarity of sentiments and extract the supporting phrases from the tweet.
- Employed pre-trained model Global Vectors (GloVe) for word embeddings of 27K sentences to capture semantics in sentences.

Lock the Vote | BlockChain | Spyder | MySQL DB | Postman | Flask
- Designed an electronic voting system using BlockChain technology and MySQL Database to make the voting process more secure transparent and reliable.
- Demonstrated more than 8 voting steps using Flask framework and Postman APIs.