LinkedIn
paljain@iu.edu

Palak Jain

Website

Indianapolis, IN
(765)-775-3286

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: 4.0/4.0)** **January 2021 - Ongoing**
Bachelor of Engineering, Electronics and Telecommunication, Devi Ahilya University, India, **(GPA: 7.2/10.0)** **July 2014 - May 2018**

Experience

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%).

Systems Engineer **January 2019 - November 2020**
Infosys Limited **Hyderabad, India**

- Organized 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.

Data Analyst Intern **July 2018 - December 2018**
Virtuosity **New Delhi, India**

- Compiled and studied a large amount of data of $\sim$ 10K customers to understand trends in sales for different geographical zones.
- Proposed localized inventory of popular products resulting in delivery time reduction by 40% with 10% increment in profit margin.
- Created interactive reports with PowerBI to develop focused marketing strategies for the regional customer base, improved ROI by 25% on digital marketing expenditure.

Technical Skills

Languages	Python, R, Core Java, HTML
Database Management	MySQL, SQL Server, Hadoop, Informatica, MS Access
Analytics Tools	PowerBI, Tableau, MS Excel
Statistical Skills	Data Cleaning, Data Visualization, Clustering, Regression, Exploratory Data Analysis, Neural Networks, NLP, Word Embedding, Dimension Reduction, Hypothesis Testing. (Libraries: Numpy, Pandas, Sklearn, NLTK, Gensim, Spacy)

Projects

Diabetes Onset Prediction | NLP | Pytorch | Artificial Neural Networks (ANN) | Jupyter Notebook | 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%.

Digit Recognizer | Computer Vision | Tensorflow | Keras | Convolution Neural Networks (CNN) | Google Colab

- Implemented CNN model for hand-written digit recognition using MNIST dataset with 60K training and testing digits.
- Achieved accuracy of 98% in classification of hand-written digits using Sequential API.

Tweet Sentiment Extraction | NLP | Keras | Word Embeddings | Artificial Neural Networks (ANN)

- Created ANN model with Functional API to analyze polarity of sentiments and extract the supporting phrases from the tweet.
- Utilized pre-trained model Global Vectors (GloVe) for word embeddings of 27K sentences to capture semantics in sentences.

Lock the Vote | BlockChain | Spyder | MySQL | 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.