Preet Modi

Email : prmodi@iu.edu
Linkedin: https://www.linkedin.com/in/preetjmodi/
Mobile : +18123182011
Website: https://impreetmodi.github.io/

Education

- **Indiana University Bloomington** — Bloomington, IN
  Masters in Data Science; GPA: 3.82 — *Aug 2022 - May 2024*
  Courses: Advanced Database Concepts, Big Data Management, Introduction to Statistics, Algorithms, Computer Science, Data Mining, Cloud Computing, Machine Learning, Software Engineering, Predictive Analytics (Kelley School of Business)
- **Dharmsinh Desai University** — Gujarat, India
  Bachelor of Information Technology ; GPA: 3.7 (8.29/10.0) — *Aug 2018 - May 2022*

Skills Summary

- **Languages:**: Python, R, SQL, Java, C, C++, HTML, CSS, JavaScript, C#, Linux, ReactJS, Node JS
- **Database & Tools:**: SQL Server, PostgreSQL, Hive, MongoDB, Tableau, PowerBI, Airflow, Kafka, SAP, SAS, Excel, VS Code, AWS, GCP, PySpark, Databricks, Snowflake, Git, Azure, EC2, MATLAB
- **Data Science:**: ETL, Predictive Modeling, Regression, Classification Trees, Time Series Analysis, Data Warehousing, Natural Language Processing, Hypothesis Testing, Artificial Intelligence, Statistical Analysis, Data Visualization

Experience

- **Indiana University** — Bloomington, IN
  Graduate Research & Teaching Assistant — *August 2023 - Current*
  - **SAS, SQL, MS Power Tools, Data Visualization, Carnegie Classification**: Collaborating with Dr. Victor Borden, I engaged in data metric analysis, processing datasets of over 1 million records, and developed 10 novel interactive visualizations for Carnegie Classification.
  - **Python, Flask, Kubernetes, NoSQL, Cloud Computing, Big Query, Distributed Computing, Docker**: As a Graduate Teaching Assistant for INFO-I 535 Management, Access, And Use of Big And Complex Data, I crafted 13 assignments, graded for 80 students, achieving 95% completion and boosting student engagement by 30%.
- **Sacoma Specialty Products,LLC** — Edinburgh, IN
  Data Engineering Intern — *May 2023 - August 2023*
  - **SQL, Epicor, SAP, Excel**: Integrated Epicor and SAP systems with AWS services, utilizing custom Business Activity Queries (BAQs), resulting in a 20% improvement in supply chain efficiency.
  - **Amazon Redshift, QuickSight, HPC**: Created a centralized data lake, streamlining data extraction, transformation, and loading (ETL) processes, reducing data processing time by 40% and enhancing data accessibility.
- **Indiana University** — Bloomington, IN
  Data Analyst — *Oct 2022 - May 2023*
  - **Power BI, Advanced Excel, Dash, R, Tableau & DAX**: Worked with Residential Program and Services to conduct financial data analysis for housing and dining facilities, utilizing DAX (Data Analysis Expressions) for tailored computations, leading to a 15% decrease in operational expenses and a 20% uptick in revenue.

Academic Projects

- **Topic Modeling on Customer Reviews - Yelp.com**: This Yelp.com review analysis project, driven by Python and tools like **LDA, Scikit-learn, and Power BI**, aimed to enhance user experience and support businesses. It involved data collection (web scraping on Yelp's API) using **Selenium** and text preprocessing **(spaCy, NLTK, Gensim)**. An ensemble machine learning model classified reviews, while **Latent Dirichlet Allocation (LDA)** uncovered topics, visualized with **Matplotlib and Power BI**. The results provided actionable recommendations for improving customer satisfaction, loyalty, and business performance. (May 2023)
- **Exploratory Data Analysis for Bureau of Transportation Statistics Flight Performance**: Implemented a robust **data pipeline**, devised a storage model in a **NoSQL** server, and executed an algorithm using a parallel programming framework, **Hadoop**, resulting in a 30% reduction in data processing time. Proposed and implemented a cleaning improvement solution, explored a big data cloud platform environment, and developed a reliable data management plan. Leveraged the **K-Means Clustering** algorithm to enhance data analysis accuracy, achieving an impressive **90%** precision rate. Additionally, conducted statistical analysis using **R** and **Stata**. (Aug 2022)
- **Claim Severity Prediction using Computer Vision and Machine Learning**: Designed and implemented a state-of-the-art machine learning model utilizing **Convolutional Neural Networks (CNNs)** and a suite of Computer Vision libraries, including **OpenCV, TensorFlow, and Detectron2**, Meta AI's platform for object detection and segmentation. This model accurately predicted auto insurance claim severity based on images of damaged vehicles, achieving an outstanding **95%** accuracy rate in distinguishing repairable from total loss cases. (Feb 2023)

Publications

- **"Insurance Management with Premium Prediction ", Volume 9, Issue XII, International Journal for Research in Applied Science and Engineering Technology (IJRASET) Page No: 1222-1238, ISSN: 2321-9653 (Impact Factor: 7.429)**: - DOI: https://doi.org/10.22214/ijraset.2021.39416
- **"An efficient Artificial Neural Network for Coronary Heart Disease Prediction ", Volume 9, Issue XII, International Journal for Research in Applied Science and Engineering Technology (IJRASET) Page No: 1474-1483, ISSN: 2321-9653 (Impact Factor: 7.429)**: - DOI: https://doi.org/10.22214/ijraset.2021.39559