Preet Modi

Linkedin: https://www.linkedin.com/in/preetjmodi/ Email : prmodi@iu.edu
Website: https://impreetmodi.github.io/ Mobile: +1 8123182011

Education

- **Indiana University Bloomington** Bloomington, IN
  *Masters in Data Science; GPA: **3.84*** *Aug 2022 - May 2024*
  Courses: Advanced Database Concepts, Big Data Management, Data Structures & Algorithms, Statistics, 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*

Experience

- **Research Data Scientist** August 2023 - Current
  Indiana University *Bloomington, IN*
  - **SAS, SQL, MS Power Tools, Data Visualization, Virtualization, Cassandra, SPSS**: Collaborating with Dr.Victor Borden, 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, Information Systems, BigQuery, Distributed Computing, PL/SQL**: As a Graduate Teaching Assistant for INFO-I 535 Management, Access, And Use of Big And Complex Data, Crafted 13 assignments, graded for 80 students, achieving 95% completion and boosting student engagement by 30%.
- **Data Science Intern** May 2023 - August 2023
  Sacoma Specialty Products,LLC *Edinburgh, IN*
  - **SQL, Epicor, SAP, CRM, Pyspark**: 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, DataIQ**: Created a centralized data lake, streamlining data extraction, business analytics, and loading (ETL) processes, reducing data processing time by 40% and enhancing data security.
- **Data Analyst** Oct 2022 - May 2023
  Indiana University *Bloomington, IN*
  - **Power BI, Advanced Excel, Dash, R, Tableau & DAX**: Collaborated with IU Residential Program and Services for financial analysis for housing and dining facilities, utilizing Power BI for data analytics and data reporting. Conducted A/B testing and multivariate analysis for web-app optimization, leading to a 15% increase in revenue.

Academic Projects

- **Epicor-Driven Data Enlightenment: Transforming Business Intelligence**: Implemented a data-driven dashboard to enhance operational efficiency and decision-making. Integrated data from diverse sources and conducted analyses to extract insights, refining data manipulation skills. Utilized **Amazon Redshift** for data pipelines and **Excel Macros** for streamlined processing. Developed interactive dashboards using **Power BI, Tableau & Quicksight**, and in ERP systems like **Epicor** and **SAP**. Provided stakeholders with real-time visibility into key indicators, facilitating informed decisions. Wrote **Business Activity Queries (BAQ)** in Epicor and generated **MES** dashboards, achieving a **40%** increase in efficiency.(Aug 2023)
- **HPC Analytics Dashboard Application Development**: Engineered an application for High Performance Computing (HPC) with a theoretical peak performance of 1 Petaflop (PF) and an infrastructure comprising over 10,000 CPU cores and 44 GPU cards. Integrated **Oracle RDMS** for data storage and management. Leveraged **Pandas, Numpy**, **Matplotlib**, and **Seaborn Python** libraries and **ReactJS** for data visualization. This project entailed implementing Software Development Life Cycle methods such as **Agile** and **Scrum**, utilizing **Jira** for project management. Harnessing problem-solving, communication skills, documentation, and cross-functional abilities, the analysis delivered actionable business insights.(Nov 2023)
- **Topic Modeling on Credit Card Fraud Detection**: Implemented a robust **data pipeline** using **Python** and **ETL tools i.e Alteryx**; employed **DataIQ** for efficient data ingestion and preprocessing; collected banking transaction data from financial APIs using **Selenium** and preprocessed it with **spaCy, NLTK, Gensim**; developed an ensemble machine learning model with **XGBoost** to classify transactions and enhance credit card fraud detection, achieving an F1-score of 90%, while **Latent Dirichlet Allocation (LDA)** uncovered topics, visualized with **Matplotlib, Plotly**; Using the **OpenAI GPT-3 API & Flask** integrated a Chatbot to provide interactive fraud-related insights to improve fraud prevention. (Feb 2024)

Skills Summary

- **Languages**: Python, R, SQL, Java, C, C++, HTML, CSS, JavaScript, C#, Linux, ReactJS, Node JS
- **Database & Tools**: SQL Server, PostgreSQL, Hive, MongoDB, Power BI, Hadoop, Kafka, SAP, SAS, TensorFlow, Keras, Looker, Airflow, AWS, GCP, Salesforce, Spark, Databricks, Snowflake, Github, SSIS, SSRS, Qlik, Azure, EC2, MATLAB
- **Data Science**: ELT, BI, Predictive Modeling, Regression, Classification Trees, Time Series Analysis, Data Warehouse, Natural Language Processing, Hypothesis Testing, Artificial Intelligence, Statistical Analysis, Data Architecture, GIS

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