Milan Prajapati

British Columbia, Canada • +1-7788551756 • mbprajapati1732@gmail.com • LinkedIn • Portfolio • GitHub

SUMMARY

A results-driven and client-focused data scientist with 2 years of experience in data analytics, machine learning, statistical modeling, and software development. I have honed my skills in programming languages such as Python, R, and JavaScript, as well as data visualization and reporting tools like Tableau and Microsoft Power BI. I have corporate experience along with having worked on several projects related to credit score analysis and prediction, advertisement campaigns, database management, and statistical analysis. With a problem-solving mindset and strong organizational skills, I am confident in my ability to deliver results in a data-driven world.

EXPERIENCE

Data Entry   October 2022 - Current
Save On Foods, Vancouver, BC.
- Conducting Extract, Transform, and Load (ETL) processes to extract data from various sources, transform it to fit business needs, and load it into a target database. Assisting with data cleansing, sorting, and formatting tasks to ensure data quality and integrity.
- Demonstrated proficiency in data entry and data verification by accurately inputting data into computer systems or databases while ensuring the completeness and accuracy of the data entered.
- Conducting regular backups and updates to ensure the integrity and security of data, and implementing measures to maintain data quality and consistency.

Instructional and Teaching Assistant   January 2023 - April 2023
Northeastern University, Vancouver, BC.
- Offered assistance to learners by explaining fundamental principles of statistical modeling and prediction modeling, as well as instructing on the use of Python and R programming languages.
- Assisted the professor in data collection, organization, and statistical analysis for the thesis using appropriate tools. Developed visualization techniques such as graphs and charts to effectively communicate the results.
- Reviewed and edited the drafts of the thesis and papers for accuracy, completeness, and quality.

Data Scientist   July 2021 - September 2022
Pandavaz, Gujarat, India
- Utilized advanced statistical methodologies, such as probability distributions and inferential statistics, to identify insights from given information utilizing the R programming language.
- Developed interactive dashboards using PowerBI, Tableau, and Excel to visualize and present relevant information extracted from the backend database.
- Designed and managed databases for 17 websites and 4 web applications using database management systems such as MySQL, Oracle, and Firebase, ensuring their optimal performance and security.
- Created various prediction models using forward and backward selection techniques, which involved selecting the most significant predictors based on statistical metrics such as p-values, t-values, and adjusted R-squared.
- Created models were then evaluated for their accuracy using statistical methods such as ANOVA, AIC, BIC, and other metrics. The results were presented in a report that included visual evidence of the model's performance.
- Utilized Power BI and Tableau to create insightful visualizations of trends and predictions derived from dataset analysis or simulated data. Explained the process step by step and engaged with stakeholders through visual presentations to instill confidence in the research.

Data Scientist Intern   July 2020 - June 2021
Pandavas, Gujarat, India
- Conducted exploratory data analysis to identify correlations and dependencies among various columns of the dataset and provided insights into how those relationships impact the overall data structure.
- Performed advanced data analysis by applying statistical methods including regression analysis, clustering, and machine learning algorithms to uncover patterns and trends in the dataset.

- Assisted with data translation by converting data from one format to another as per business requirements.
- Conducted detailed analysis of business processes and document findings for the purpose of streamlining operations and improving efficiency. Facilitated data translation across various departments and provided technical support to ensure seamless customer service.

PROJECTS

Credit Score Analysis and Prediction
- Acquired and extracted data from three distinct sources by utilizing Python programming language and Microsoft Excel. Conducted data cleaning and transformation tasks to prepare the data for further analysis. Loaded the formatted data into a centralized database for storage and management.
- Performed statistical analysis on banking data to identify variables affecting credit scores, resulting in 93% of the elements being accurately computed.
- Created 6 distinct predictive models utilizing the forward and backward model selection technique, aimed at identifying the most relevant predictors for the given data set.
- Evaluated model performance with 82% accuracy using 70% of the dataset through diagnostic analysis.
- Performed residual analysis on the models using a 30% dataset to obtain a residual value of 3%.

Earning Analysis of San Francisco
- Conducted exploratory data analysis to identify key aspects of the job market and performed Monte Carlo simulations to predict possible outcomes of uncertain events and situations in the job market.
- Utilized exploratory data analysis techniques to identify and classify the interrelationships between various columns of the job market dataset, thereby determining their impact on the job market trends and outcomes.
- Formulated and executed multiple hypothesis tests to evaluate various scenarios and determine their statistical significance and conclusions.
- Produced a detailed report featuring easily understandable visualizations, charts, and graphs generated using Excel and Tableau, and provided an extensive explanation of the work performed using customer-centric and user-friendly language.

TECHNICAL SKILL

Languages and DataBase: Python, R, JavaScript, MySQL, Oracle, SQL Server, HTML, CSS, ReactJS, MongoDB
Package: Scikit-learn, Pandas, NumPy, SpaCy, Matplotlib, TensorFlow, Beautifulsoup, ggplot2, Shiny
Data Visualization and Reporting: Microsoft Power BI, Microsoft Excel, Tableau, Microsoft Word, Microsoft Office, PowerPoint
Statistics and Analytics: Hypothesis Testing, Regression Models, Prediction Models, Probability distribution, Data simulation
Tools and Technologies: R Studio, Visual Studio code, GitHub, Pycharm, Jupyter Notebook, Git
Organizational Skills: Leadership, Project Management, Critical Thinking, Problem-Solving

CERTIFICATION AND ACHIEVEMENTS

- **Certifications:** IBM Data Science - Coursera, Learn MySQL – Udemy
- **Achievements:** HackerRank Python Gold Badge, HackerRank SQL Gold Badge

EDUCATION

Masters Of Professional Studies (M.S), Data Analytics   March 2024
Northeastern University, Vancouver, BC
- Deep Learning | Data Modelling | Data Science | Machine Learning and AI |Natural Language Processing

Bachelor of Engineering (B.E.), Computer Engineering
Gujarat Technological University, Ahmedabad, Gujarat   July 2021
- SQL | DBMS | Data Structure | Algorithms | Data gathering |Algorithms | Data Computation | Probability