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 - June 2023

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.

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.
- Implemented natural language processing techniques, including various linguistic units, to encode customer reviews on the client's website. Utilized advanced algorithms to accurately identify the sentiment of each review, distinguishing between positive and negative feedback.
- 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.
- Developed interactive dashboards using PowerBI, Tableau, and Excel to visualize and present relevant information extracted from the backend database.
- 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.

Full Stack Developer
July 2020 - June 2021

Pandavas, Gujarat, India

- Managed and maintained a portfolio of 12 websites for diverse product-based companies, leveraging ReactJS in frontend, Django in Backend, and SQL database, resulting in efficient and dynamic web applications with seamless user experience and high client satisfaction.
- Build reusable components using React (Next JS), TypeScript, HTML, CSS, and Styled Components, ensuring consistency across products.
- Implemented API integration to extract live data from web sources using API calls and secure credentials, optimizing data collection processes for improved accuracy and timeliness.
- Collaborated with cross-functional teams, including strategy, release management, engineering, validation, and program management, to validate assumptions and provide data-driven recommendations.
- Performed comprehensive website maintenance tasks for multiple websites, including bug identification and resolution, functionality testing, and performance optimization, ensuring seamless user experience and enhanced overall website performance.

EXPERIENTIAL LEARNING AND FREELANCING PROJECTS

Artificial Intelligence Generated Resumes data analysis | Sponsor: CoverQuick

- Proficiently converted unstructured AI-generated resume data from JSON format into structured dataframes by leveraging advanced techniques for data extraction and enhancing data usability.
- Conducted data cleansing of big data, transformed and anonymized country names to a standardized format, enabling accurate identification of optimal client base for business sponsorship.
- Leveraged natural language processing techniques to analyze resumes and provide highly relevant job recommendations based on individual qualifications, skills, and experience.
- Utilized resume data, including experience and graduation details, to accurately infer the age of individuals by applying Python PySpark, Yake, and data algorithmic, contributing to comprehensive candidate profiling and demographic insights.
- Generated comprehensive detailed reports revealing hidden patterns within the database, effectively addressing stakeholder queries and concerns by providing robust evidence-based solutions, with visualization generated using Power BI facilitating informed decision-making and strategic planning.

Match-making Automation and Data Analysis | Sponsor: Qunuby

- Analyzed existing data collection methods and business structure to assess the current workflow, enabling identification and recommendation of suitable automation and digitalization strategies.
- Developed a new optimized database schema to efficiently store data for further processing, and successfully executed the migration of legacy data into the new database, ensuring seamless data transfer and enhanced data management capabilities.
- 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%.

TECHNICAL SKILL

Languages and DataBase: Python, R, JavaScript, MySQL, Oracle, SQL Server, HTML, CSS, ReactJS, MongoDB
Package: Scikit-learn, Pandas, NumPy, SpaCy, PySpark, 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