**Teja Bairaboina**

**Data Scientist / AI Engineer**

**Plano, TX ‬| +1 2142998373 |** **Rishitha.23g@gmail.com**

**PROFESSIONAL SUMMARY:**

* 4+ years of experience in Proactively identified and mitigated projects ensuring smooth progression through each phase of the SDLC.
* Spearheaded the development and implementation of predictive models and AI algorithms, employing a diverse array of programming languages including Python, R, JavaScript, Go, SQL, Scala, and SAS, to drive data-driven decision-making processes across various business functions.
* Designed and executed complex data analysis and processing tasks using advanced data science packages such as Matplotlib, NumPy, SciPy, Pandas, MATLAB, and ggplot2, enhancing data quality and insight extraction.
* Mastered the use of visualization tools including Matplotlib, Tableau, Power BI, Minitab, and MS Excel with Pivot Charts and Macros, to develop intuitive dashboards and reports that significantly improved business intelligence and stakeholder communication.
* Advanced the application of Machine Learning and NLP using tools and libraries like Scikit Learn, RapidMiner, Data Robot, NLTK, spaCy, GPT-3, BERT, and Hugging Face's Transformers, contributing to the development of sophisticated models that accurately interpret and predict data patterns.
* Leveraged cloud database tools and platforms such as Azure, AWS, and GCP to architect and maintain scalable, high-performance data storage solutions, ensuring data integrity and accessibility.
* Implemented cloud-based tools and Deep Learning frameworks including Big ML, Google Analytics, Azure AutoML, TensorFlow, and PyTorch to facilitate the creation and deployment of state-of-the-art AI models.
* Utilized multipurpose data science tools and Big Data technologies like Apache Spark, SAS, KNIME, Apache Hadoop, Spark, Hive, and Pig to process and analyze large datasets, extracting actionable insights that informed strategic decisions.
* Conducted research on the latest LLM advancements, contributing to internal innovation projects.
* Managed ETL processes and data integration tasks using SSIS, ensuring the reliable transfer and transformation of data across systems, while utilizing SSRS for comprehensive reporting solutions.
* Maintained high standards of data governance and version control using GIT, SVN, and Bitbucket, enhancing team collaboration and project tracking in a fast-paced development environment.
* Continuously engaged in learning and integrating emerging technologies and methodologies in DS and AI, contributing to the organization's leadership in innovation and competitive advantage.
* Monitored LLM-based applications, using feedback for continuous model improvement and relevancy.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| **Methodologies** | Agile, Waterfall |
| **Programing languages** | Python, R, Java script, GO, SQL, Scala, SAS |
| **Packages** | Matplotlib, NumPy, SciPy, Pandas, MATLAB, ggplot2 |
| **Visualization Tools** | Matplotlib, Tableau, Power BI, Minitab, MS Excel, Pivot charts, Macros |
| **Machine Learning Algorithms and NLP Tools** | KNN, KMN, SVM, Gradient Boosting, Random Forest, Scikit Learn RapidMiner, Data Robot, NLP |
| **Cloud Database tools** | Azure, AWS, GCP |
| **Cloud-based tools** | Big ML, Google Analytics, Azure automated machine learning (AutoML) |
| **Natural Language Processing (NLP) Libraries** | NLTK,LLM, spaCy, GPT-3, BERT, Hugging Face's Transformers |
| **Deep Learning Frameworks** | TensorFlow, PyTorch |
| **Multipurpose Data science Tools** | Apache Spark, SAS, KNIME, Apache Flink, Alteryx |
| **Big Data Tools** | Apache Hadoop, Hadoop, Spark, Hive, Pig |
| **Reporting and other Tools** | SSIS, SSRS |
| **Data Bases** | MongoDB, My SQL |
| **Version control** | GIT, SVN, Bitbucket |

**EDUCATION:**

* Bachelors in Computer Science | Jawaharlal Nehru Technological University Hyderabad- 2020
* Master in Information Technology Management | Concordia University St Paul - 2023

**PROFESSIONAL EXPERIENCE:**

**Client: Microsoft Corporation, WA Sep 2023 – Present**

**Role: Associate Data Scientist**

**Project Description:**

In a team of four, we utilized Agile methodology to simplify complex concepts into user-friendly experiences and products. I built a predictive model using decision trees to forecast boat failures, involving comprehensive data collection and preprocessing across 10+ unstructured datasets. I applied text mining to analyze survey data on boat covers, uncovering customer pain points and design insights. Using various regression techniques, I developed a Python tool to predict a boat's top speed and acceleration based on engine parameters. Additionally, I integrated LLM into products and developed a Python API to fetch feedback from MongoDB, updating Power BI dashboards.

**Responsibilities:**

* Working in a team environment of 4 peoples with Agile methodology, transformed complex concepts into simple, elegant, and approachable experiences, services, and products.
* Built a predictive model using decision trees to predict the failure of a boat within next 10 trips, form the stage of data collection and pre-processing such as assessment of data quality, new feature creation, feature selection and exploratory analysis on 10+ unstructured data sets and provided recommendations for more efficient data capturing and storage.
* Analyzed survey data on boat covers using Text Mining Principles such as topic modeling and sentiment analysis to discover pain points for customers and provided insights that aid in better design of boat covers.
* Performed feature selection and compared various regression techniques such as Linear, Lasso, Ridge, Elastic Net and SVR to create a tool in Python that can predict the top speed and maximum acceleration possible for a boat using the engine parameters and operating conditions.
* Collaborated with teams to integrate LLM into products, providing technical guidance on their capabilities.
* Developed a python API code to fetch feedback data from MongoDB and update Power-BI dashboards.

**Environment:** Python, scikit-learn, Pandas, NumPy, NLTK, MongoDB, Power BI, Decision Trees, Regression Techniques, Text Mining, LLM Integration

**Client: Carelon, VA July 2022 – Aug 2023**

**Role: Data Scientist/Software Developer**

**Project Description:**

In an Agile development environment, I collaborated with senior developers to implement voice recognition using Google’s Speech-to-Text for multilingual support and integrated Dialogflow to enhance natural language understanding. I developed KPI dashboards with Tableau to provide insights to stakeholders and optimized LLM performance through hyperparameter tuning. I addressed imbalanced fraud datasets with sampling techniques and cost-sensitive algorithms in Python, implemented webhook services for dynamic conversations, and worked with cross-functional teams to align AI strategies with business goals, improving customer satisfaction by 20%.

**Responsibilities:**

* Utilized the Agile methodology of development for Requirements, Planning, Design, and Deployment.
* Worked with senior developers to implement voice recognition features using Google's Speech-to-Text for a multi-lingual support system.
* Assisted in integrating the Dialogflow platform into the existing customer service framework to enhance natural language understanding capabilities.
* Derived insights from datasets and developed KPI dashboards to present them to stakeholders using Tableau.
* Applied optimization and hyperparameter tuning to balance LLM performance and computational efficiency.
* Collected data needs and requirements by Interacting with the other departments.
* Tackled highly imbalanced Fraud dataset using under sampling with ensemble methods, oversampling with
* SMOTE and cost sensitive algorithms with Python Scikit-learn.
* Implemented webhook services for dynamic conversational experiences, integrating with backend systems to provide real-time data and personalized user interactions.
* Regularly collaborated with cross-functional teams to align conversational AI strategies with business goals, contributing to a 20% increase in overall customer satisfaction ratings.

**Environment:** Agile methodology, Google Speech-to-Text, Dialogflow, Tableau, LLM hyperparameter tuning, Python Scikit-learn, SMOTE, ensemble methods, webhook services, and real-time data integration to enhance multilingual support and AI-driven insights.

**Client: PB Tech PVT Ltd June 2020 – March 2022**

**Role: Data Scientist**

**Project Description:**

Developed customized applications to address business challenges in the Telecom domain, including critical prediction and decision automation. Improved call routing efficiency by 30% with a new IVR solution. Conducted data analysis using Python, SAS, and R, and performed exploratory data analysis with libraries such as NumPy, SciPy, and pandas. Deployed NLP applications with LLMs and TensorFlow, managed configurations with Chef on GCP, and applied various algorithms and tools for well production optimization. Utilized R Studio and SAS Text Miner for data visualization and customer behavior analysis.

**Responsibilities:**

* Created customized applications to make critical predictions, automate reasoning and decisions for business problems across Telecom Domain for various clients.
* Analyzed complex data and identified anomalies, trends and risks to provide useful insights to improve internal controls.
* Played a key role in designing and deploying an IVR solution that improved call routing efficiency by 30%, resulting in reduced customer wait times.
* Performed data analysis using python libraries like NumPy, SciPy, pandas and matplotlib.
* Perform exploratory data analysis with descriptive and inferential statistics in Python, SAS, R, and data imputation in Scikit-learn, Pandas, and Natural Language Tool Kit (NLTK).
* Developed and deployed NLP applications using LLMs, such as chatbots and content generation tools.
* Apply TensorFlow in numerous hands-on activities using the command line. Also Using TensorFlow on real-world data sets including photos and text
* Used Chef for configuration management of hosted Instances within GCP. Configuring and Networking of Virtual
* Private Cloud (VPC).
* Run Different set of algorithms to maximize well production (KNN, KMN, SVM, Gradient Boosting, Random Forest...etc.) using different tools and libraries: RapidMiner, Scikit-learn, Genix...etc.
* Worked on R Studio plotting the graphs based on survey data collected and used SAS Text miner to understand the customer behavior.
* Participated in features engineering such as feature intersection generating, feature normalize and label encoding with Scikit-learn preprocessing.

**Environment:** Python (NumPy, SciPy, pandas, matplotlib), SAS, R, Scikit-learn, NLTK, TensorFlow, Chef, GCP, VPC, RapidMiner, Genix, R Studio, SAS Text Miner.