**Saivijay Panuganti**

**Email:** [**panugantisaivijay@gmail.com**](mailto:panugantisaivijay@gmail.com) **Mobile:** +1 (571) 464-9829

**Objective**

Seeking a Data Scientist position to leverage my expertise in data analysis, machine learning, and statistical analysis to drive business decisions and create innovative solutions in such a way that I can add value to the organization and become a better professional myself.

**Summary**

* As a Data Scientist and Data Analyst with over 7 years of experience in the financial, marketing and telecom sectors with a strong background in business, math/statistics/analysis and programming, I excel at utilizing data science methodologies to address complex business challenges.
* My expertise lies in data analysis, data visualization, machine learning, and natural language processing.
* Strong SQL skills for data mining, data manipulation, extraction, aggregation, and analytics in multiple database technologies.
* Collaborated with Data Science teams to develop data engineering solutions using Python, PySpark, SQL, Hive, AWS, and Azure.
* Created reproducible data visualizations with Tableau, Power BI, and Python to examine trends and anomalies.
* Aided in data modeling by generating source data from existing tables.
* Successfully implemented machine learning models for predicting high-risk clients, benefiting the business department's pilot operation.
* Employed ML techniques to clean and analyze text/case data from different surveys to enhance customer experience.
* Highly comfortable in working with cloud provider platforms such as AWS, GCP and Azure, especially AWS.
* Highly proficient in using visualization tools such as matplotlib, ggplot2, Tableau, and Power BI to design dashboards.
* Built Pareto chart, Market Basket analysis, Waterfall chart, Bump chart and stories using Tableau.
* Participated in several client meetings to showcase different analytical and visualization capabilities, prepared weekly status report, and shared with the management.
* Performed exploratory data analysis like calculation of descriptive statistics, detection of outliers, assumptions testing, factor analysis, etc., in Python.
* Designed data visualizations, technical documentation, and non-technical presentation materials to communicate complex ideas and solutions to business partners.
* As a self-motivated and goal-oriented Data Scientist, I am passionate about leveraging data to solve tangible, real-world problems and am confident in my ability to make a substantial impact on your team.

**Skills**

* Programming: Python, Visual Basic
* Machine Learning: Python (Scikit-Learn, statsmodels, tensorflow, Pandas, NumPy), pyspark.ml
* Visualization: Tableau, Python (matplotlib, seaborn, plotly, ggplot2), PowerBI
* Database: SQL Server, MySQL, NoSQL
* Others: AWS, Notebooks, EDA, Excel (Macros, Pivot tables, Slicers), MS Office, Databricks, PySpark, Statistical analysis, CAPM, Market Basket analysis, Pareto chart.

**Work Experience**

**Data Scientist**

*Capital One, Richmond, VA, USA*

*January 2022 - June 2023*

**Description:** The objective was to partner with a cross-functional team of data scientists, software engineers, machine learning engineers and product managers to build machine learning models to improve the customer experience and develop readable reports and dashboards to the non-technical teams.

**Roles & Responsibilities:**

* Responsible for working with data engineers to building the ML models and performing EDA to get useful insights from the data.
* Built regression models such as SVR, XGboost to predict Customer Lifetime Value.
* Built classification models such as Logistic Regression, SVM, Decision Tree, Random Forest to predict Customer Churn Rate.
* Used F-Score, AUC/ROC, Confusion Matrix, MAE, RMSE to evaluate different Model performance.
* Built pipelines, performed One Hot encoding (OHE) and hyperparameter tuning to improve the performance of the models and get the best performing parameters.
* Use advance Excel tools to clean and prepare data for analysis on Tableau, and create Pivot tables, slicers, data lookup etc.
* Performed time series analysis and forecasting by leveraging ARIMA, Naive and Smoothing models.
* Performed Data Imputation by establishing threshold using Scikit-Learn in Python.
* Performed Exploratory data analysis on how the customers interact with the money to make personalized marketing for the customer demographics.
* Worked with the distributed data/ computing tools such as AWS, Databricks, PySpark, and snowflake.
* Leveraged the capabilities of Spark for managing the large customer base.
* Used Machine learning models to perform the fraud detection.
* Leveraged a broad stack of technologies such as Python, Tableau, PowerBI, AWS to reveal the insights hidden within huge volumes of numeric and textual data.
* Built ML foundation models through all phases of development, from design through training, evaluation, validation, and implementation.
* Showcased my interpersonal skills to translate the complexity of my work into tangible business goals in an understandable way for both technical and non-technical teams.

**Environment:** Python (Matplotlib, Seaborn, NumPy, Pandas, Scikit-Learn), Machine Learning, Tableau, Power BI, Excel, Anaconda, Jupyter notebooks, AWS, PySpark, Snowflake.

**Data Scientist**

*Fannie Mae, Reston, VA, USA*

*July 2020 - December 2021*

**Description:** The objective was to collaborate with product owners and platform teams to understand the business needs to enhance the delivery of business applications by implement new modeling applications to support risk measurement, financial valuation, decision making, and business performance. Analyzed surveys from the lenders using NLP.

**Roles & Responsibilities:**

* Responsible for creating data visualizations, technical documentation, and non-technical presentation materials to communicate complex ideas and solutions to business partners.
* Maintained and queried the databases using MySQL and AWS Redshift to the reporting teams.
* Setup Tableau Dashboard reporting tool and explained the reporting teams on the Visualization and dashboard capabilities.
* Extracted MyServices ticketing data into SQL database and generated interactive Tableau dashboards for Incident and Problem tickets.
* Pulled data and built charts, dashboards as needed using Python, Excel, Tableau and create graphical representation of the information which is easily understandable for others.
* Created documentations along with visualizations and effectively communicated the insights to the respective teams.
* Responsible for data preprocessing, Univariate and Bivariate analysis, Exploratory data analysis on the extracted data using Pandas and Numpy python libraries.
* Performed data cleansing activities including feature engineering, missing value imputation, outlier removal, variable transformation, Matplotlib and Seaborn python libraries and created visualization in the Jupyter notebooks to deliver monthly SLA reports for the executives.
* Analyzed the data to identify trends or pattens that are appearing and how they are impacting on the stability of the customer base.
* Created ML models using Linear Regression, SVM, Decision tree algorithms for predicting the number of loans that can be repurchased. This helped the Business teams to plan for the staffing of Quality Control reviewers.
* Created ML models using Decision Trees, Random Forest, AdaBoost, and XGBoost algorithms to identify the important features for the loans with more defects and the loans being repurchased. This helped the Underwriters to focus on these important features when reviewing the loans.
* Created NLP models using BeautifulSoup, STOPWords, Lemmatizer and WordNet Python libraries to analyze the surveys submitted by the lenders during the review process.
* Created clusters using Unsupervised learning algorithms like K-means and DBSCAN to identify the production tickets opened by the Customer Support teams to avoid the assignment of the tickets to the wrong groups due to insufficient knowledge.

**Environment:** Python (Matplotlib, Seaborn, NumPy, SciPy, Pandas, Scikit-Learn), MySQL, Machine Learning, Tableau, NLP, Power BI, AWS (Redshift, S3), Excel, Anaconda, PySpark

**Data Analyst**

*Middleware Information Technology Services Pvt Ltd, Hyderabad, TG, India*

*March 2018 - June 2020*

**Description:** The objective was to predict the probability of churn rate of customers so the company could redirect their marketing campaigns towards these customers to reengage these customers with products with the goal of identifying behavioral patterns of the customers that lead to disengagement

with the product.

**Roles & Responsibilities:**

* Collaborated with Data Engineering department to implement the Extract-Transform-Load Pipelines, wrote and optimized queries in SQL to perform Data Extraction to fit the analytical requirements.
* Performed Exploratory Data Analysis by observing customer specific features by using Matplotlib, Seaborn in Python and dashboards in Tableau.
* Performed missing values imputation, anomaly detection and dimensionality reduction.
* Participated in Feature Engineering such as Feature Generation, Feature Normalization and Label Encoding with Scikit-Learn in Python.
* Used Correlation Analysis to estimate the strength of association between various predictors and with the response.
* Used NumPy, SciPy, and Pandas in Python for Data Manipulation and Data Analysis to prepare the data for predictive modeling.
* Reported statistical and analytical insights to executives for effective strategic positioning in the marketplace.
* Implemented Predictive Modeling for Classification using Logistic Regression, Decision Trees, and Support Vector Machines.
* Experimented with Ensemble Learning methods such as Bagging, Random Forests and Boosting with Gradient Boosting for model optimization.
* Applied Performance Metrics based on Accuracy, Precision, Recall, F-Score, Receiver Operating Characteristics and Area Under Curve to evaluate the model.
* Summarized results with rich data visualizations and presented them as a dashboard using Tableau.

**Environment:** MySQL, Python (Matplotlib, Seaborn, NumPy, SciPy, Pandas, Scikit-Learn), Machine Learning (Logistic Regression, Decision Trees, Bagging, Random Forests, Boosting, Support Vector Machines), Tableau, Extract-Transform-Load Pipelines

**Data Analyst**

*Tikona Infinet Pvt Ltd- Noida, UP, India*

*March 2016 - February 2018*

**Description:** The primary objective was to analyze telecom data to better understand customer behavior, network performance, and market trends, ultimately aiding in improved network connectivity, customer satisfaction, and business efficiency.

**Roles & Responsibilities:**

* Analyzed and processed customer data for dataset preparation to help implement marketing and recommendations strategies.
* Built a customer-centric machine learning algorithm for the retention of customers in the retail business of the company.
* Regression models to understand the in-store stock capacity of the Retail outlets.
* Ingested customer data into HIVE tables from HDFS (Hadoop Distributed File Systems) for customer segmentation.
* Maintained and queried the database using MySQL.
* Investigated the data quality issues and assist Data engineers to correct them.
* Implemented Customer Segmentation based on RFM principles by data partitioning to enable targeted marketing.
* Worked on demographics, and collaborative filtering-based recommendation strategies to recommend suitable financial services best fitting to the customer needs.
* Generated Reports and Visualizations using visualization tool by pulling the data from ERP.
* Used advance Excel tools to clean and prepare data for analysis on Tableau, and create Pivot tables, slicers, data lookup etc.
* Monitored the services requested to us on the daily basis to discover any trends and patterns.
* Performed Data cleaning and Exploratory Data Analysis on our services data to assist my team.

**Environment:** Python (Matplotlib, Seaborn, NumPy, Pandas, Scikit-Learn), MySQL, Hadoop, Hive, Tableau, Excel, Machine Learning (Regression)

**Education:**

**University of Maryland, Baltimore County (UMBC), MD, USA**

* **Program:** Master’s in data science.
* **Certification:** Data Science post baccalaureate certification.

**MallaReddy College of Engineering, TG, India**

* **Program:** B. Tech in Computer Science and Engineering.