**Gayatri**

[**gayatri.v0524@gmail.com**](mailto:gayatri.v0524@gmail.com)

**956-616-5012**

**PROFESSIONAL SUMMARY:**

* 9 years of experience in Data Analytics, Machine Learning, Data Mining, Predictive modelling, Data Visualization, Data Acquisition, and Data Validation with large datasets of Structured and Unstructured data.
* Extensive experience in Text Analytics, developing different Statistical Machine Learning Algorithms, Data Mining solutions to various business problems and generating Data visualizations using R, Python and Tableau.
* Professional working experience in Machine Learning algorithms such as Linear Regression, Logistic Regression, Naive Bayesa, Decision Trees, K-Means Clustering and Association Rules.
* Strong Data Analysis skills using business intelligence, SQL and MS Office Tools.
* Proficient in SAS, MATLAB, and hands-on experience in writing queries in SQL (Teradata, MySQL) and Informatica to extract, transform and load (ETL) and Visualization using Tableau
* Proficient in statistical analysis and hypothesis testing to derive meaningful insights from data.
* Developed high-performance data processing applications in Rust, leveraging its memory safety and low-level control features to optimize data analysis workflows.
* Experience in working on different Databases/Data warehouses like Teradata, Oracle, AWS Redshift, Snowflake
* Good Understanding of Data ingestion, Airflow Operators for Data Orchestration, and other related python libraries.
* Extensive experience in developing various real - time dashboards and reports using different Data visualization tools like Qlik Sense, Tableau, Power BI and Power pivot etc.
* Have Extensive Experience in IT data analytics projects, Hands on experience in migrating on premise ETLs to Google Cloud Platform (GCP) using cloud native tools such as BIG query, Cloud Data Proc, Google Cloud Storage, and Composer.
* Expertise in Cluster Analysis, Principal Component Analysis (PCA), Association Rules, Recommender Systems.
* Experience with toolset GIT, Jenkins, Unix scripting.
* Developed multiple dashboards using Tableau and PowerBI as BI tools.
* Migrated on premise ETLs to Google Cloud Platform (GCP) using cloud native tools such as BIG query.
* Involved in Analysis of Programs and JCL’s
* Implement ETL process in Alteryx to extract data from multiple sources (SQL Server, XML, Excel, CSV) and schedule workflows.
* Familiar with predictive models using numeric and classification prediction algorithms like support vector machines and neural networks, and ensemble methods like bagging, boosting and random forest to improve the efficiency of the predictive model.
* Worked on Text Mining and Sentimental analysis for extracting the unstructured data from various social Media platforms like Facebook, Twitter, and Reddit.
* Hands on experience with R Studio for doing data pre-processing and building machine learning algorithms on different datasets.
* Coordinated with team and Developed framework to generate Daily adhoc reports and Extracts from enterprise data from BigQuery.
* Excellent knowledge in Data Analysis, Data Validation, Data Cleansing, Data Verification and identifying data mismatch.
* Performed data analysis and data profiling using complex SQL on various sources systems including Oracle.
* Experience with Data visualization using tools like Ggplot, Matplotlib, Seaborn, Tableau and using Tableau software to publish and present dashboards, storyline on web and desktop platforms.
* Experienced in Data manipulation for loading and extraction with python libraries such as NumPy, SciPy and Pandas for data analysis and numerical computations.
* Well experienced in Normalization, De-Normalization and Standardization techniques for optimal performance in relational and dimensional database environments.
* Excellent in creating various artifacts for projects, which include specification documents, data mapping and data analysis documents.
* Excellent SQL programming skills and developed Stored Procedures, Triggers, Functions, Packages using SQL, PL/SQL.
* Skilled in Advanced Regression Modeling, Correlation, Multivariate Analysis, Model Building, Business Intelligence tools and application of Statistical Concepts.
* Expertise in Data Integration, Data Cleaning, Data Analysis, and Profiling, and Data Import and Export using multiple ETL tools such as SQL Server, SSIS, and SAS.
* Perform Data Analysis on the analytical data present in AWS S3, AWS Redshift, Snowflake, and Teradata using SQL, Python, Spark, and Data bricks.
* Very keen in knowing newer techno stack that Google Cloud platform (GCP) adds.
* Extensive data analysis/data reporting/system analysis experience with experience working in data processing environment using ETL tools/technologies.
* Extensive hands-on experience and high proficiency with structures, semi-structured and unstructured data, using a broad range of data science programming languages and big data tools including R, Python, Spark, SQL.
* Extensive experience in applying various statistical techniques, including regression analysis, ANOVA, and clustering algorithms.
* Collaborated with the lead Data Architect to model the Data warehouse in accordance with FSLDM subject areas, 3NF format, and Snowflake schema.
* Flexible with Unix/Linux and Windows Environments, working with Operating Systems like Ubuntu13/14.

**DATA SCIENCE EXPERTISE:**

Data and Quantitative Analysis \* Exploratory Data Analysis \* Model Building \* Decision Analytics \* Predictive Modeling \* Big Data Queries and Interpretation \* Data Mining and Visualization Tools \* Natural Language Processing \* Machine Learning Algorithms \* Time Series Analysis \* Research, Reports and Forecasts \* Computer Vision \* Deep Learning \* Data-Driven Personalization.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| **Languages** | Python, R, SQL |
| **Machine Learning Methods:** | Linear regression, Logistic regression, Decision tree, RandomForest, K nearest neighbor, K mean, NLP, ARIMA, APRIORI, ETS, SVM |
| **Packages:** | Pandas, NumPy, SciPy, Scikit-learn, Stats models, NLTK, Plotly, Matplotlib, Seaborn, plyR, dplyR, data.table and sqldf, tidyR, Reshape2 |
| **Cloud Technologies:** | Amazon Web Services (AWS), Microsoft Azure (familiar), Amazon EC2, Google Cloud Platform (GCP) |
| **Databases** | SQL, Hive, Impala, Pig, Spark SQL, Databases SQL-Server, MySQL, MS Access, HDFS, HBase, Teradata, Netezza, Mongo DB, Cassandra, SAP HANA, Tera Data |
| **Reporting Tools** | MS Office (Word/Excel/Power Point/ Visio), Tableau, Power BI, Crystal reports XI, Business Intelligence, SSRS, Business Objects 5.x/ 6.x, Cognos7.0/6.0. |
| **Version Control Tools** | GitHub. |
| **Project Methodologies** | Agile, Scrum |
| **BI Tools** | Tableau, Tableau server, Tableau Reader, Power BI, SAP Business Objects, OBIEE, QlikView, SAP Business Intelligence, Amazon Redshift, and Azure Data Warehouse |
| **Operating System** | Windows, Linux, Unix, Macintosh HD |

**PROFESIONAL EXPERIENCE:**

**CLIENT: Wells Fargo Bank, Charlotte, NC** **Feb 2021 to Till Date**

**Role: Data Scientist**

**Responsibilities:**

* Initial project was on loan application pre approval using Supervised and Unsupervised learning.
* The second project was using Computer Vision for analyzing the customer body language. Using only customers’ faces and behavioral characteristics, banks will authorize transactions.
* Familiarity with Bayesian statistics and its application in decision-making processes.
* Human behavior detection based on tracking trajectory analysis and pattern recognition.
* Applied data-mining, machine learning and/or graph analysis techniques for a variety of modeling and relevance problems involving users and their interests in various content types
* Design, develop, implement and execute marketing campaigns for US card customers using Unica Affinium campaign, Snowflake, AWS S3, Pyspark, Databricks.
* Created one-time bulk data movement through Mainframe JCL and Sqoop from RDBMS to HDFS.
* Utilized Rust for data preprocessing, cleaning, and transformation tasks, ensuring data quality and efficiency.
* Ensured regulatory compliance by implementing robust data privacy and security measures, including anonymization techniques and encryption protocols.
* Configured Azure platform offerings for web applications, business intelligence using Power BI, Azure Data Factory, etc.
* Experience in the field of Data Analytics and Machine learning Algorithms with good knowledge and hands-on experience in **R, Python and SAS EMINER and DATAIKU platform.**
* Created database class using Python for lift and shift cloud projects. Invoked Python scripts in airflow.
* Extensive experience in Data Visualization including producing tables, listings, graphs using various procedures and tools such as Tableau.
* Enmeshed in Publishing the reports from Power BI desktop to POWER BI services, Workspace, Dashboard and App creation in POWERBI SERVICES
* Manage end to end complex data migration, conversion, and data modeling (using Alteryx, SQL), and create visualization using tableau to develop high quality dashboards.
* Developing executive summary, departmental and provider independent dashboards using SAP business objects and later migrating it into Qlik Sense
* Mapped data types of Big query and Oracle in Python database class.
* Performed transfer learning to identify human emotion and detect faces.
* Used OpenCV SSD to detect faces, and used Face Net, which has 1024 features to recognize and compare faces.
* Proficient in version control using Git for tracking changes, collaborating with team members, and managing project history.
* Used NumPy for data distribution techniques and Pandas for data cleaning.
* Performed Exploratory Data analysis (EDA) to further explore the dataset, detect the outliers and extract important variables both graphically and numerically.
* Performed feature scaling (Standardization/standard scaling) and Min. Max scaling (Normalization).
* Created custom Rust libraries and modules for specific data science tasks, promoting code reusability and maintainability.
* Can work parallelly in both GCP and Azure Clouds coherently.
* Applied resampling methods like Synthetic Minority over Sampling Technique (SMOTE) to balance the classes in large data sets.
* Implemented algorithms such as Principal Component Analysis (PCA) for dimensionality reduction and for normalizing large datasets.
* Implemented classification algorithms such as Logistic Regression, KNN and Random Forests to predict the customer credibility based on previous records and provide discretion to the marketing team about eligibility.
* Analyzed the customer behavior based on interaction using Bayesian inference, GRU, LSTMS as part of Sentiment Analysis.
* Write complex SQL scripts to analyze data present in different Databases/Datawarehouse’s like Snowflake, Teradata, and Redshift.
* Performed dimensionality reduction using Principal Component Analysis (PCA) and for Word Embeddings TF-IDF and Word2Vec and Doc2Vec.
* Deployed models as python package, as api for backend integration and as services in a micro-services architecture with a kubernetes orchestration layer for the dockers containers
* Strong foundation in multivariate statistics for analyzing relationships among multiple variables.
* Improved accuracy of predictive models using Support Vector Machines and Random Forests with XG Boost.
* Build data pipelines in airflow in GCP for ETL related jobs using different airflow operators.
* Used hyperparameter-tuning methods like Bayesian optimization, random search and grid search for model tuning.
* Processed CSV files using Python Programming and loaded data to Big Query from GCS bucket.
* Experience with Git workflows such as feature branching, pull requests, and code reviews to ensure code quality and collaboration.
* Built machine-learning models and constructed multilayer perceptions for Deep Neural Networks (DNN) to identify fraudulent applications for loan pre-approvals and to identify fraudulent credit card transactions by comparing the previous transactions.
* Measured the performance using Confusion matrix and Classification report.
* Used different Bagging and Boosting methods to increase the accuracy of the training model and used scores to evaluate the models like F score, AUC-ROC curves.

**Environment:** Python (Scikit-Learn/Scipy/Numpy/Pandas/Matplotlib/Seaborn), Tableau, Machine Learning algorithms (KNN, Decision Tree, Random Forest, Logistic Regression, Support Vector Machines, XGBoost), Microsoft Excel, MySQL

**Client: At & T, Plano, TX**  **Sep 2019 - Jan 2021**

**Role: Data Scientist**

**Responsibilities:**

* Participated in all phases of data mining, data cleaning, data collection, developing models, validation, and visualization.
* Accomplished data pipeline process, collected required data from different sources and converted into structured form using SQL.
* Incorporated Exploratory Data Analysis to identify the correlation between variables, multicollinearity, and hidden patterns, trends and seasonality using NumPy and Pandas libraries to perform data analysis.
* Create automated solutions using Databricks, Spark, Python, Snowflake, HTML
* Experience in moving data between GCP and Azure using Azure Data Factory
* Performed PCA, backward feature selection, correlation analysis for Dimensionality Reduction of the data to achieve the accuracy in the results.
* Performed SMOTE analysis to deal with unbalanced distribution of the data over the training data set.
* Documented and submitted reports on descriptive statistics and graphs of predictor variables
* Implemented parallelism and concurrency using Rust to improve the performance of data-intensive computations and simulations.
* Carrying out specified data processing and statistical techniques Using SAS such as sampling techniques, estimation, hypothesis testing, time series, correlation, and regression analysis Using R.
* Created dashboards, reports, visualizations and analytics using Qlik Sense followed by ETL to integrate different data subsets from 4 different databases SQL server, MySQL, Postgres and Oracle
* Implemented various classification models such as Logistic Regression, Decision Trees, Random Forest, KNN, XGBoost, and SVM and applied most efficient algorithm to predict the results.
* Migrated on premise ETLs to Google Cloud Platform (GCP) using cloud native tools such as BIG query.
* Used cloud shell SDK in GCP to configure the services Data Proc, Storage, BigQuery
* Performed K-Fold cross-validation to test models with different batches to optimize the model and prevent overfitting.
* Utilized SAS and SQL to extract data from statewide databases for analysis.
* Participated in feature engineering such as feature generating, PCA, feature normalization and label encoding with Scikit-learn preprocessing.
* Achieved 90% customer monthly retention by predicting the likelihood of returning customers using a logistic regression model in R.
* Perform data comparison between SDP(Streaming Data Platform) real time data with AWS S3 data and Snowflake data using Databricks, Spark SQL, and Python
* Designed and implemented end-to-end systems for Data Analytics and Automation, integrating custom visualization tools using R, Tableau, and Power BI.
* Integrated Rust applications with existing data science tools and frameworks, such as Python and R, for seamless data analysis pipelines.
* Improved data cleansing and mining process based on R and SQL, resulting in a 50% of time reduction.
* Conducted analysis and pattern on customer’s needs in different location, different categories and different months by using time series modeling techniques

**Environment:**Python 3.x, (Scikit-Learn/Scipy/Numpy/Pandas/Matplotlib/Seaborn), Tableau, Machine Learning algorithms (KNN, Decision Tree, Random Forest, Logistic Regression, Support Vector Machines, XGBoost), Microsoft Excel, MySQL

**Client: Allegion, Carmel, PA                                                                     Feb 2018 - Aug 2019**

**Role: Data Scientist/ Machine Learning Engineer**

**Responsibilities:**

* Built predictive machine learning, simulation, and/or statistical models using Python.
* Generated comprehensive analytical reports by running SQL queries against current databases to conduct data analysis.
* Implemented a predictive maintenance model for cable network infrastructure, reducing downtime and improving overall network reliability. The project involved the analysis of historical network performance data and the deployment of machine learning algorithms for proactive issue identification.
* Maintained and developed complex SQL queries, stored procedures, views, functions, and reports that qualify customer requirements using Microsoft SQL Server.
* Performed Data Cleaning, features scaling, features engineering using pandas and NumPy packages in python.
* Used pandas, NumPy, Seaborn, SciPy, matplotlib, sci-kit-learn in python for developing various machine learning algorithms.
* Performed Information Extraction using NLP algorithms coupled with Deep learning (ANN and CNN), Keras and TensorFlow.
* Participated in all phases of datamining; data collection, data cleaning, developing models, validation, visualization and performed Gap analysis.
* Performed data manipulation, Data preparation, Normalization and Predictive modelling. Improved efficiency and accuracy by evaluating model in Python.
* Generated the reports and visualizations based on the insights mainly using Tableau and developed dashboards for the company insight teams.
* Proficient in Predictive Modelling, Data Mining Methods, Factor Analysis, ANOVA, Hypothetical testing, normal distribution and other advanced statistical and econometric techniques.
* Performed data manipulation, Data preparation, Normalization and Predictive modelling. Improved efficiency and accuracy by evaluating model in Python.
* Generated the reports and visualizations based on the insights mainly using Tableau and developed dashboards for the company insight teams.

**Environment**: Python, R, Machine learning, deep learning, NLP, Tableau, SQL, MapReduce, Tensor flow, Oracle, pandas, NumPy, Seaborn, SciPy, matplotlib, sci-kit-learn, NLP, ANN, CNN, TensorFlow, Keras, Anova, Tableau

**Client: W3 Softech, India**  **Jun 2014 - Aug 2017**

**Role: Software Engineer**

**Responsibilities:**

* Designed, developed (using embedded C), verified and validated Engine Management System (EMS) software applications for automotive using Embedded C.
* Experienced in emission control strategy design for automotive systems.
* Developed python scripts to automate the validation process.
* Collaborated with other subsystems to design and develop streamlined software interfaces and algorithms.
* Involved in interacting with customers and system engineers to gather subsystem design requirements
* Conceptualize and present multiple ideas in design reviews with Principal Engineers, Technical Leads and other subsystem product owners
* Worked for development and maintenance of subsystem software components and systems for multi-platform Locomotive Control System using Agile and Iterative Software Development Life Cycle
* Develop prototype software based on conceptualized idea and carry out, SIL, MIL and HIL testing
* Develop test procedure and carry out Locomotive tests on production software to verify design and quality requirements.
* Implemented face detection algorithm for operator alertness detection, for LocoCAM product
* Write documentation for software requirements, design and usage guidelines.

**Environment:** Oracle, SQL, UNIX, AWS Snowball, S3 Transfer acceleration, SQL Assistant, Python, BI, DWH, Birst, Elastic Load Balancer, A/B testing, Tableau.