**FULL NAME**

**(123) 456-7890 | email@mail.com**

|  |  |  |  |
| --- | --- | --- | --- |
| |  | | --- | | **PROFESSIONAL SUMMARY** |  * Data Scientist with 8+ years of experience in telecom and supply chain domains, adept at leveraging statistical modeling, data extraction, cleaning, and exploration across both structured and unstructured datasets. Expert in large-scale machine learning (ML) and deep learning (DL) algorithms, providing insights that drive strategic decision-making processes * Proficient in Python (Pandas, NumPy) with a deep understanding of the data science life cycle and expertise in delivering robust, high-quality code for large-scale data operations * Strong experience in data visualization using Matplotlib, Seaborn, Plotly, and Tableau, creating insightful, interactive dashboards to present trends and actionable insights to stakeholders * In-depth knowledge of statistical fundamentals and advanced techniques such as ANOVA, t-tests, and dimensionality reduction (PCA, LDA), with practical applications of OOP principles * Proficient in Object-Oriented Programming (OOP), including concepts such as inheritance, polymorphism, and encapsulation, with the ability to design scalable, maintainable systems in Python * Skilled in building ML/DL models for tasks like classification, regression, and NLP, with hands-on experience using Artificial Neural Networks (ANNs), CNNs, RNNs, and LSTMs * Familiarity with cloud computing technologies like AWS (EC2, S3, SageMaker) and big data tools including Spark, PySpark, MongoDB, and Kafka for optimizing large-scale data processes * Strong team player, quick learner, and highly motivated to continuously explore new technologies and methodologies, ensuring timely delivery of data-driven solutions * Adept at causal modeling and advanced statistical analysis to identify relationships within data and designing experiments to validate findings * Developed interactive Tableau dashboards to visualize insights, creating actionable data stories and communicating key trends for stakeholders in a user-friendly format * Quick learner, continuously embracing new technologies and methodologies with a passion for personal and professional growth * Strong team player with excellent communication skills, highly motivated, self-driven, and capable of working both independently and collaboratively to meet project objectives  |  | | --- | | **TOOLS/TECHNOLOGIES** |   **Python libraries:** SQLAlchemy, Pandas, Scikit-learn, Matplotlib, Request, Selenium, NumPy, SciPy  **Technologies:** Python, Apache Spark, Java, JavaScript, HTML, CSS, Tailwind CSS, JSON, Node JS, React JS,  **Software frameworks:** Django, Flask, Pyramid, Django Rest Framework (DRF)  **ML/DL:** Random Forest, Decision Tree, CNN, RNN, LSTM, Linear/Logistic Regression, XgBoost, Clustering  **AWS** **Services**: AWS S3, SageMaker, RDS, EC2, Lambda, Bedrock  **IDE**: Jupyter Notebook, Visual Studio Code  **Databases:** MySQL, PostgreSQL, MongoDB, Redis, Cassandra, Neo4j  **CI/CD pipelines:** Git, Docker, GitLab, Jenkins, Linux, GitHub Actions  **SDLC**: Agile   |  | | --- | | **EDUCATION** |   **San Jose State University** **San Jose, CA**  *Bachelor of Science: Applied Mathematics, Concentration in Statistics*  *Minor: Computer Science*  **PROJECT EXPERIENCE** |

**Data Scientist**

*Verizon, New York*Jan 2022 – Present

* Developed system models, predictive algorithms, and solutions for prescriptive analytics problems, applying advanced data mining techniques to large datasets to drive business insights
* Analyzed business problems by using appropriate statistical models such as regression analysis, causal modeling, and leveraged Pandas and NumPy for data manipulation to generate meaningful insights and support data-driven decision-making
* Built and deployed statistical and machine learning models to solve large-scale, customer-focused problems, using statistical methods like decision trees, clustering, SVM, and neural networks to address business challenges
* Conducted preliminary and exploratory data analysis (EDA) using descriptive statistics and visualization libraries like Matplotlib and Seaborn, handled anomalies such as missing data imputation and duplicate removal to ensure data quality and integrity
* Applied machine learning algorithms including decision trees, NLP, regression models, clustering, and neural networks, using Python’s Scikit-learn package alongside Pandas and NumPy to identify key patterns and volumes, improving business processes
* Performed data cleaning and feature selection using MLlib in PySpark and worked with deep learning frameworks like TensorFlow to build Artificial Neural Networks (ANNs) and Convolutional Neural Networks (CNNs) for predictive modeling, including churn rate prediction
* Applied causal modeling to identify cause-and-effect relationships within customer behavior, enhancing model accuracy and improving stakeholder decision-making
* Built predictive models using ensemble methods such as gradient boosting and neural networks (Keras) to forecast outcomes like sales amounts, enhancing business forecasting accuracy
* Performed feature engineering techniques including label encoding, PCA, and feature normalization using Scikit-learn, Pandas, and NumPy for model optimization and improving predictive accuracy
* Collaborated with data engineers and operational teams to implement ETL processes, writing and optimizing SQL queries for data extraction, transformation, and loading to fit analytical and reporting needs
* Leveraged python for efficient data manipulation and created robust data models to ensure the accuracy and reliability of insights and utilized python for efficient data manipulation and created robust data models to ensure the accuracy and reliability of insights
* Utilized Object-Oriented Programming (OOP) principles in Python to design reusable code structures, implementing classes and methods for better organization of complex data analysis workflows, enhancing maintainability and scalability of projects
* Collaborated with cross-functional teams to communicate insights, explain statistical concepts in clear terms, and drive data-informed decisions across the organization
* Developed interactive dashboards and visualizations using Tableau to communicate complex data-driven insights, ensuring stakeholders and decision-makers can access and interpret information easily
* Applied Tableau for building data visualizations that highlight key trends and patterns from complex datasets, improving data-driven decision-making across various business functions

**Data Scientist**

*Publix, Lakeland* Jul 2019 – Nov 2021

* Responsible for clarifying business objective, data collection, data wrangling, data processing, machine learning modeling, model tunning, deploying models
* Worked with cross-functional teams to understand the data requirements and optimize problem-solving skills to find the necessary data to support business problems
* Leveraged Pandas and NumPy for efficient data manipulation including handling missing values and performing data aggregations to deliver meaningful insights from the data and employed OOP concepts to strucutre and encapsulate data processing methods effectively
* Utilized Pandas DataFrames to merge and join datasets effectively, ensuring a comprehensive view of customer behavior and performance metrics for enhanced analysis
* Performed cluster analysis to access sales agent performance and gain insights by performing segmentation tasks that group agents with similar performance metrics together using Python packages
* Applied regression analysis to optimize call volume by examining the relationship between call volume and influencing factors to predict and refine the call volume recommendations that help to reduce the abandoned call rate
* Assessed the potential risk associated with different customer segments and used statistical modeling and regression analysis to model and analyze the risk score with different independent variables and conducted various statistical analysis including ANOVA and many hypotheses testing methods
* Visualized complex datasets and communicated insights using Tableau to build interactive dashboards for stakeholders, enabling data-driven decision-making
* Created Tableau reports to effectively monitor key performance indicators (KPIs) and provide ad-hoc visual analysis to explore trends and patterns
* Designed Tableau dashboards for executives, allowing for seamless tracking of operational performance and enabling high-level data visualization to support strategy discussions
* Created database schemas to support data integrity and optimize storage solutions, facilitating effective data management for analytics and modeling tasks
* Evaluated model performance using techniques like R square, adjusted R square, confusion matrix, AUC-ROC curve
* Communicate insights with third-party partners such as marketing agencies and CRM platforms to target and personalize outreach strategies
* Collaborate with data engineering team to extract customers data from various sources and assured data quality and data integrity and optimized data collection procedures on a weekly and monthly basis using Python
* Collaborated with the data scientist team and BA team to analyze on building a predictive model based on the requirements using various ML algorithms
* Implemented Pandas, NumPy, Seaborn, SciPy, Matplotlib, Scikit-Learn in Python to perform data cleaning and data visualization, processing techniques like checking the skewness and distributing the data normally by log transformation and Box-Cox

**Data Scientist**

*Ulta Beauty, Bolingbrook*Sep 2017 – Jul 2019

* Worked independently and collaboratively throughout the analytical lifecycle including data extraction, data preparation, design and implementation of scalable analysis and solutions, and documentation of results
* Collaborated with technical/non-technical resources across the business to support and integrate our efforts and supported data management workflow from data collection, storage, analysis to training and validation
* Testing multiple classification model like random forest, SVM, logistic regression and gradient boosting. Also performed hyper-parameter tuning on the models to optimize the model predicting power in Python
* Experimented various DL algorithms and ensured that the model has low false positive rate
* Obtained knowledge of image processing algorithm: encoding/decoding, feature detection and matching, image segmentation and transformation
* Used a combination or various filter sizes for convolutions, max pooling, selection of activation functions (ReLu and SoftMax), dropout functions and batch normalization to regularize data
* Built a CNN model and NLP techniques to predict category to product using the label (title) of the image and modified the model to accommodate the CNN of the text blocks and integrated them both (image model and text model) to produce a reliable and highly accurate model
* Extensively used Databricks clusters and PySpark to extract and analyze text data with NLP techniques such as word-embedding and word similarities
* Involved in all jobs for pipelining the data from the databases through to analysis and reporting
* Simulated the model multiple times changing the hyperparameters such as learning rates, epochs, batch size, activation function, number of hidden layers and units, dropouts, and initial weights
* Developed machine learning/deep learning algorithms to classify the behavioral patterns using Machine Learning models such as random forests and decision trees on IDE like Vscode Studio, Pycharm.
* Developed machine learning strategies for risk analysis using multiple regression
* Worked with DevOps team to deploy applications on AWS EC2 and supported in automation testing
* Utilized AWS SageMaker to develop, train, and deploy machine learning models for efficiency and scalability of model training and tunning within the data pipeline
* Wrote complex SQL queries to extract and manipulate data from relational databases to ensure accurate and reliable data analysis
* Developed data models using SQL to ensure data integrity and optimize performance in data retrieval, supporting the machine learning pipeline and reporting efforts
* Created stored procedures with SQL to automate data extraction processes, enhancing efficiency in retrieving data for machine learning and analytical tasks

**Data Scientist**

*Glossier, New York*May 2016 – Aug 2017

* Responsible for clarifying business objective, data collection, data wrangling, data processing, machine learning modeling, model tunning, deploying models
* Worked closely with internal stakeholders such as business teams, product managers, engineering teams, and partners
* Collaborated with the data scientist team and BA team to analyze on building a predictive model based on the requirements using various ML algorithms
* Implemented Pandas, NumPy, Seaborn, SciPy, Matplotlib, Scikit-Learn in Python to perform data cleaning, processing techniques like checking the skewness and distributing the data normally by log transformation and Box-Cox
* Implemented various data processing techniques to manipulate the unstructured, and structured data and tackled highly imbalanced data set under sampling and over-sampling techniques like SMOTE
* Used Spark MLlib to leverage the computational power of Spark for data processing and building machine learning models to improve the performance and optimization for the large-scale dataset using Spark Context, Spark SQL, and Spark Data Frame
* Wrote PySpark queries to clean, impute and manipulate over 100 million records of customer data for EDA and modeling
* Connected the data sources such as AWS RDS and S3 to import data into QuickSight for data visualization and analyzing trends and summary statistics
* Built a multi-text classifier on the business data glossary to classify more than 3000 attributes using NLTK, Word2Vec to build word embedding
* Worked intensively on AWS Services like SageMaker, Lambda, EC2, EMR, and S3 for large-scale data storage, processing, and automation, to optimize resource usage for machine learning tasks
* Involved in performing text analytics with customer feedback in emails and call transcripts, built a text classifier and Sentiment Analyzer using Recurrent Network LSTM and a huge word cloud in Python
* Leverage AWS SageMaker to build, train, tune, and deploy state of art machine learning models and deep learning models
* Performed routine maintenance of data pipelines and machine learning models to address system errors and optimized workflows for efficiency and performance
* Created detailed and interactive reports for stakeholders by generating data visualizations with AWS QuickSight and performance summaries to present insights on model performance, resource utilization, and business impact for informed decision-making