Vamsi Krishna

**Data Scientist**

rishitha.23g@gmail.com

+1 2142998373

Texas

# PROFESSIONAL SUMMARY

* Results-driven Data Scientist with 10+ years of experience in leveraging data analytics to drive business decisions and solve complex problems.
* Worked in the entire data science project life cycle and actively involved in all the phases including data extraction, data cleaning, statistical modelling, and data visualization with large data sets of structured and unstructured data.
* Extensive experience in Text Analytics, developing different Statistical Machine Learning, Data Mining solutions to various business problems and generating data visualizations using R, Python.
* Extensive expertise in developing end-to-end machine learning pipelines, from data collection and preprocessing to model training, evaluation, and deployment in production environments.
* Skilled in implementing state-of-the-art machine learning algorithms including deep learning models (CNNs, RNNs, GANs), ensemble methods, and reinforcement learning techniques.
* Proficient in data mining tools like R, SAS, Python, SQL, Excel, ecosystems Staff leadership and development
* Experienced with Machine Learning Algorithm such as Logistic Regression, KNN, SVM, Random Forest, Neural Network, Linear Regression, Lasso Regression and K-Means.
* Creating from scratch Machine Learning and NLP solutions for Big Data on top of Spark using Python.
* Experienced in developing and deploying machine learning models in production environments using frameworks like TensorFlow, PyTorch, or scikit-learn.
* Expertise in Deep Learning, Predictive Model Development, and Machine Learning. Proficient in utilizing R, Python, and Git to develop and train accurate machine learning models.
* Experienced in deploying machine learning models using containerization technologies like Docker and orchestration tools such as Kubernetes for scalable and reliable production deployments.
* Extensive experience in collaborating with cross-functional teams to understand business requirements, formulate data-driven strategies, and deliver innovative solutions.
* Proficient in utilizing cloud computing platforms such as AWS, Azure for scalable data processing and model deployment.
* Experienced in working with structured and unstructured data sources including relational databases, APIs, and streaming data for real-time analytics.
* Skilled in designing and implementing A/B tests and experimental methodologies to measure the impact of data-driven initiatives and optimize business outcomes.
* Strong analytical and problem-solving skills with a keen attention to detail and a passion for continuous learning and professional development.
* Familiarity with big data technologies such as Hadoop, Spark, and Kafka for distributed computing and real-time data processing.
* Proficient in developing predictive models and algorithms, utilizing statistical analysis and machine learning techniques.

# SKILLS

|  |  |
| --- | --- |
| Languages:  | Python, PySpark SQL, Scala**,** R, SAS |
| Databases: | MySQL, PostgreSQL, DB2, Oracle, MongoDB |
| ML Techniques: | Linear/Logistic Regression, KNN, Support Vector Machines, Decisions Tree, Random Forest, Neural Networks, Principal Component Analysis |
| Visualization Tools: | Power Bi, Tableau, Looker, Alteryx |
| Big Data & Cloud Computing | Spark, Hadoop, Azure (Data factory, Azure SQL, Cosmos DB, Databricks, Data Lake Storage, Containers), AWS (S3, EMR, Athena, SageMaker) |
| Mathematical Optimization:  | Probability theory, Linear Algebra, Calculus, Time series analysis |
| DevOps:  | MLflow, Azure DevOps, Git, Docker, Kubernetes, Jenkins |
| Other Skills: | Scikit-learn, TensorFlow, PyTorch, Keras, Pandas, NumPy |

# EDUCATION

**Master’s in business Analytics |** University Of Texas Arlington, Texas

**Bachelor’s in electronic and communication |** Gandhi Institute of Technology, Visakhapatnam, India

# PROFESSIONAL EXPERIENCE

**Client: State Farm, Dallas, TX** Dec 2022 – Present

 **Role: Sr. Data Scientist**

**Description:**

Experienced Data Scientist proficient in designing and implementing MLOps workflows to maximize revenue opportunities. Skilled in analyzing large data sets using Python, pandas, and NumPy to derive insights and drive business decisions. Developed and deployed machine learning models using TensorFlow, PyTorch, and scikit-learn. Utilized Docker, Kubernetes, and ML Flow for containerization and model management. Built predictive models for customer segmentation and forecasting using techniques like XGBoost and random forests. Created dashboards and reports with Power BI to communicate insights. Expertise in SQL, NLP, web scraping, and distributed computing with Hadoop.

**Responsibilities:**

* Designed and implemented workflows for efficient data flow to maximize revenue opportunities using MLOps
* Analyzed and interpreted large and complex data sets using Python, pandas, and NumPy, deriving meaningful insights to drive business decisions.
* Utilized Docker and containerization techniques to package and deploy machine learning models and applications.
* Extensively worked with Kubernetes and container orchestration platforms to manage and scale machine learning deployments.
* Developed machine learning solutions using Python, incorporating libraries like TensorFlow, PyTorch, or scikit-learn.
* Led the design and implementation of a robust and scalable MLOps platform, facilitating seamless ML model deployment, monitoring, and orchestration.
* Utilized SQL, and Python to create customer segments and improve revenue, responsiveness, and conversions using Market mix modeling techniques.
* Implemented ML Flow projects to package ML code, making it reproducible and shareable.
* Utilized ML Flow experiment tracking to log and compare model performance metrics during training.
* Built ML model for Customer Segmentation based on a member’s purchase behavior and frequency patterns.
* Partnered with model team in building the Mix Models to quantitatively estimate the effectiveness of market elements.
* Forecasted the sales and margin and built dashboards to identify missed opportunities for stakeholders.
* Performed NLP based tokenization, lemmatization, vectorization and developed NLP models for Topic Extraction, Sentiment Analysis. Analyzed and interpreted large and complex data sets using Python, pandas, and NumPy, deriving meaningful insights to drive business decisions.
* Built predictive models using techniques such as multivariate regression, random forests, and XGBoost to forecast customer behavior and improve targeting and segmentation.
* Developed and implemented web-scraping techniques to gather data from various online sources, ensuring data quality and accuracy.
* Performed topic modeling, sentiment analysis and used word embeddings for clustering.
* Collaborated with business partners for data acquisition & validation and built automated data pipelines.
* Built and maintained forecasting models to accurately predict demand and optimize inventory levels, reducing stockouts and excess inventory.
* Tracked visitor interaction and traffic on the firm websites using Google Analytics and automated reports for tracking online traffic KPI’s in Microsoft Power BI.
* Created visually compelling dashboards, reports, and presentations to communicate insights and recommendations to stakeholders.
* Built models using Statistical techniques like Bayesian HMM and ML classification models like XG Boost, SVM and, Random Forest.
* Piping and processing massive data-streams in distributed computing environments such as Hadoop to facilitate analysis (ETL).
* Assisted in conducting A/B testing across multiple channels that drive improvements in retention and risk factors experience.

**Environments:** Python, Docker, Kubernetes, ML Flow, SQL, Power BI, TensorFlow, PyTorch, scikit-learn, Hadoop, Google Analytics.

**Client: Liberty Mutual, TX** Sept 2021– Nov 2022

**Role: Data Scientist**

**Description:**

 Skilled data scientist with expertise in insurance and logistics, leveraging machine learning to enhance accuracy, reduce costs, and drive strategic decision-making. Proficient in Python, SQL, NLP, and advanced analytics.

**Responsibilities:**

* Developed and deployed machine learning models to predict insurance claim probabilities, improving accuracy by 53% over previous methods.
* Implemented algorithms to assess and mitigate risk factors, resulting in more precise underwriting decisions and reduced loss ratios.
* Utilized advanced anomaly detection techniques to identify fraudulent insurance claims, saving the company $7 million annually.
* Employed clustering algorithms to segment policyholders based on behavior and demographics, enabling targeted marketing strategies and personalized customer experiences.
* Built predictive models to forecast customer churn, enabling proactive retention strategies and increasing customer lifetime value by 48%.
* Conducted time series analysis to forecast insurance demand and optimize resource allocation, reducing operational costs by 25%.
* Implemented NLP techniques to analyze unstructured text data from customer feedback and claims descriptions, extracting valuable insights for process improvement. Played a critical role in developing complex algorithms and analytics models to solve critical business problems for clients.
* Successfully built and installed Python packages to support ML workflows and model development.
* Played a key role in creating and maintaining ML services on Unix servers, enabling seamless data processing and model deployment.
* Developed machine learning solutions using Python, incorporating libraries like TensorFlow, PyTorch, or scikit-learn.
* Leveraged SQL, Excel, and PowerBI to optimize freight costs and increase profitability by identifying opportunities to lower shipping costs.
* Utilized SQL, and Python to create customer segments and improve revenue, responsiveness, and conversions.
* Automated data extraction and cleansing processes using Python and SQL, reducing manual effort by 30% and improving data quality for analytics and reporting purposes.
* Collaborated with cross-functional teams to define business problems and design analytical solutions, translating business requirements into data-driven approaches.
* Built and maintained forecasting models to accurately predict demand and optimize inventory levels, minimizing stockouts and excess inventory costs.
* Conducted exploratory data analysis on customer behavior patterns and market trends to identify new business opportunities and optimize product assortment for an e-commerce platform.
* Provided actionable insights to senior management through data-driven presentations and dashboards, enabling informed decision-making and strategic planning across the supply chain network.

**Environments:**

Worked in insurance and logistics industries, utilizing Python (TensorFlow, PyTorch, scikit-learn), SQL, Unix servers, Excel, PowerBI, and NLP for data-driven insights and operational efficiencies.

**Client: ICICI Bank, Hyderabad (India)** July 2018 **–** July 2021

**Role: Data Scientist**

**Description:**

Experienced data scientist adept in credit risk assessment and marketing optimization, driving significant improvements in default rates, customer engagement, and conversion rates. Skilled in ML algorithms, EDA, and collaboration with stakeholders.

**Responsibilities:**

* Developed predictive models for credit risk assessment, resulting in a 15% reduction in default rates.
* Utilized different ML algorithms like Logistic regression, Gradient boosting, and neural networks to find the credit risk.
* Implemented machine learning algorithms to optimize marketing campaigns, leading to a 20% increase in customer engagement and a 10% improvement in conversion rates.
* Conducted exploratory data analysis to identify trends and patterns in financial transactions, enabling proactive fraud detection and prevention measures.
* Implemented initial Exploratory Data Analysis (EDA) and generate statistical reports like Box Plot, Scatter Chart and Heat Map using Python NumPy, Seaborn, Pandas and Matplotlib.
* Collaborated with business analysts and product managers to define project objectives, requirements, and success criteria.
* Presented findings and recommendations to senior management and key stakeholders to drive data-driven decision-making processes.
* Designed and developed algorithms for credit scoring and loan approval processes, resulting in a 25% reduction in processing time and improved accuracy.
* Analyzed customer segmentation data to personalize product offerings and enhance customer satisfaction and loyalty.
* Conducted A/B testing and statistical analysis to evaluate the effectiveness of pricing strategies and promotional campaigns.
* Collaborated with IT teams to implement data pipelines and infrastructure improvements to support scalable data analytics solutions.
* Provided mentorship and guidance to junior team members, fostering a culture of continuous learning and professional development.
* Conducted ad-hoc analysis of transactional data to identify patterns of suspicious activities and mitigate potential risks.
* Developed and maintained dashboards and reports to monitor key performance indicators and track business metrics.
* Provided actionable insights to senior management to support strategic decision-making processes.
* Mentored junior analysts and provided technical guidance on data analysis techniques, programming languages, and best practices in data visualization.

**Environments:**

Focused on credit risk assessment and marketing optimization, utilizing Python (NumPy, Seaborn, Pandas, Matplotlib), statistical analysis, A/B testing, and collaboration with IT teams for scalable solutions.

**Client: Value Labs, Hyderabad (India)** April 2016 **–** June 2018

**Role: Sr Data Analyst**

**Description:**

Dynamic data scientist with expertise in banking analytics, specializing in AHP initiatives, predictive modeling, and data-driven decision-making. Proven track record in optimizing processes and enhancing portfolio performance through advanced analytics.

**Responsibilities:**

* Spearheaded AHP initiatives to optimize decision-making processes across various banking functions, including risk management, marketing strategies, and customer segmentation.
* Built predictive models using techniques such as multivariate regression, random forests, and XGBoost to forecast customer behavior and improve targeting and segmentation.
* Designed and executed experiments to gather data, performed data cleansing and preprocessing, and applied statistical analysis and machine learning algorithms.
* Led cross-functional teams in collecting, analyzing, and interpreting large-scale datasets to identify key trends and actionable insights.
* Conducted data exploration, performed feature engineering, and developed predictive models to forecast the risk.
* Developed predictive models using advanced machine learning algorithms to forecast customer behavior, credit risk, and market trends, resulting in a significant reduction in loan defaults and improved portfolio performance.
* Implemented automated data pipelines and data cleansing techniques to ensure data accuracy and integrity, resulting in streamlined workflows and enhanced productivity.
* Collaborated with stakeholders to design interactive dashboards and data visualization tools using Python, R, and Tableau, enabling real-time monitoring of key performance indicators and business metrics.
* Conducted in-depth statistical analysis and hypothesis testing to evaluate the effectiveness of marketing campaigns and product offerings for banking clients.
* Designed and implemented customized data-driven solutions to address specific business challenges, such as customer churn prediction, cross-selling opportunities, and fraud detection.
* Presented findings and actionable recommendations to senior management and key stakeholders, facilitating data-driven decision-making processes and strategic planning initiatives.

**Environments:**

Focused on banking analytics, utilizing techniques like AHP, multivariate regression, random forests, and XGBoost, along with Python, R, Tableau, and collaboration with stakeholders for real-time monitoring and strategic planning.

**Client: DXC Technology, Chennai (India)** July 2013 **–** March 2016

**Role: Data Analyst**

**Description:**

Detail-oriented data analyst proficient in data mining, modeling, and visualization. Experienced in Agile environments, SQL, Python, Tableau, and Power BI for data-driven decision-making and process optimization.

**Responsibilities:**

* Develop a data set process for data mining and data modeling and recommend ways to improve data quality, efficiency, and reliability.
* Developed a market-basket model that provided data-driven recommendations, leading to a 70% improvement in overall sales volume.
* Working in an Agile environment, with the ability to accommodate and test the newly proposed changes at any point of time during the release.
* Analyzing data using SQL and MS Excel to identify and monitor patterns in the transnational history data.
* Import the customer data into Python using panda’s libraries and performed data analysis - found patterns in data which helped to make key decisions for the company.
* Created dashboards in Tableau and Power Bi for data visualization and present the stories to users.
* Assisted in the collection, cleaning, and validation of large datasets for analysis.
* Conducted exploratory data analysis to uncover insights and identify data quality issues.
* Developed automated scripts in Python to streamline data processing tasks and improve efficiency.
* Assisted in the preparation of reports and presentations for internal and external stakeholders.
* Participated in team meetings and brainstorming sessions to discuss data analysis approaches and findings.

**Environments:**

Worked in Agile settings, utilizing SQL, MS Excel, Python (pandas), Tableau, and Power BI for data analysis, visualization, and process improvement.