**DATA SCIENTIST**

**Rishabh**

**PROFESSIONAL SUMMARY:**

* Experienced Data Scientist with over 8+ years in data analytics, machine learning, and advanced data systems. Skilled in data visualization, statistical analysis, and complex database management, with a focus on leveraging data insights for business strategy and customer engagement. Strong expertise in Power BI, Snowflake, and SQL with a proven ability to support marketing teams in translating data into actionable insights that drive strategic decision-making.
* Expertise in Python (3.5/2.7) with libraries such as NumPy, Pandas, Matplotlib, NLTK, and Scikit-learn. Implemented various machine learning algorithms including logistic regression, random forest, XGBoost, KNN, SVM, and neural networks.
* Extensive experience in curating large datasets from diverse sources, designing feature stores optimized for analytics-ready data, and implementing graph embeddings for comprehensive data analysis.
* Skilled in advanced machine learning techniques, including graph-based machine learning and knowledge graph development to uncover patterns in complex networks and support decision-making in real-world applications..
* Designed and architected chatbots on Kore.ai platform
* Deep understanding of enterprise technologies, including Compute, Storage, Network, and Mainframe systems, to drive data modeling and optimize algorithm adjustments for system performance and reliability.
* Proficient in integrating various ML technologies and platforms, including TensorFlow, PyTorch, and scikit-learn, to work with a range of machine learning models such as vision, video, translation, and NLP.
* Experienced in Deep Learning methodologies, including CNNs, RNNs, and GANs, with applications in image recognition, NLP, and generative modeling. Skilled in NLP techniques such as word embeddings, sentiment analysis, and text generation using TensorFlow, PyTorch, and Keras. Applied AI algorithms for recommendation systems, anomaly detection, and reinforcement learning.
* Developed and deployed deep learning models for tasks such as image classification, object detection, and text summarization. Proficient in leveraging pre-trained language models such as BERT, GPT, and XLNet for transfer learning. Familiarity with advanced AI techniques such as attention mechanisms and transformer architectures.
* Experienced in designing and implementing scalable AI solutions using cloud services such as AWS, GCP, or Azure. Developed custom machine learning models and algorithms deployed on cloud infrastructure to address business objectives.
* Proficient in leveraging cloud-based AI services like Amazon SageMaker, Google AI Platform, or Azure Machine Learning, Databricks for streamlined model training and deployment. Skilled in optimizing performance and cost-effectiveness of AI applications on cloud platforms while ensuring security and compliance.
* Proficient in MLOps methodologies, including MLflow, AutoML, Jenkins, CI/CD Pipelines, Docker, and Kubernetes. Experienced in deploying machine learning models to production environments and ensuring scalability and reliability.
* Skilled in developing and managing scalable MLOps pipelines, utilizing AutoML for model quality, and implementing Vertex AI Pipelines for streamlined operations.
* Committed to maintaining data integrity, efficiency, and security while monitoring ML models in production and implementing optimizations for continuous improvement.
* Solid understanding of Big Data technologies including Spark, PySpark, Hadoop, and Hive. Experienced in SQL and NoSQL databases like MongoDB. Skilled in visualization tools such as Tableau for creating interactive dashboards. Proficient in developing API libraries and web pages using Python, Django, HTML, and AJAX.
* Passionate about leveraging data-driven insights to inform decision-making and fostering a productive team atmosphere. Strong understanding of theoretical foundations of Deep Learning, NLP, and AI. Committed to continuous learning and staying abreast of cutting-edge research and developments in the field.

**TECHNICAL SKILLS:**

**Certifications:** Deep Learning and Machine Learning Specializations, Python, SQL, R, Statistics, Google Analytics

**Programming:** Python, R, C, C++, SQL, PySpark, Java, Java, JavaScript, PHP, HTML, CSS

**Machine Learning:** Linear & Logistic Regressions, KNN, Decision Trees, SVM, Random Forests

**Ensemble Learning:** Bagging, Boosting – Adaptive, Gradient & XGBoost, Stacking, Dimensionality Reduction – PCA, Clustering & Gaussian Mixtures techniques

**Deep Learning:** ANN**,** CNN, RNN, Transformers, Autoencoders, GANs, GNNs, LSTMs, RTUs

**MLops:** MLflow, AutoML,Jenkins, CI/CD Pipelines, Docker, Kubernetes

**Libraries:** Scikit Learn, Feature Engine, Tensorflow, PyTorch, Keras, Theano, NLTK

**Analysis & Visualization Software:** Google Analytics, SAS, Tableau, STATA, Power BI, SAP Business Objects

**Statistics & Data mining:** Hypothesis Testing, Time Series Analysis, A/B testing, Association Rules

**Big Data & Cloud:** Hadoop, Spark, Kafka, Vertica, AWS, Databricks, Azure, AWS, GCP, AWS SageMaker, Azure Machine Learning, Google Cloud AI Platform, Amazon S3

**Databases:** Neo4j, Microsoft SQL Server, MySQL, MongoDB, Snowflake, Oracle

**WORK EXPERIENCE:**

# Sr. Data Scientist || Humana, Kentucky, United States || Sep 2022 – Current

# Roles & Responsibilities:

* Designed and developed interactive dashboards using **Tableau** and **Tableau Server**, enabling stakeholders to derive actionable insights.
* Leveraged **Python**, **R**, and **SAS** for advanced data extraction, statistical analysis, and machine learning to optimize healthcare services and enhance customer value delivery.
* Conducted **A/B testing** using tools like **Optimizely** and **Adobe Target**, providing data-driven recommendations for healthcare initiatives.
* Processed and analyzed structured and unstructured healthcare data, adhering to **FHIR standards** and ensuring compliance with HIPAA regulations.
* Collaborated with physicians and clinical teams to evaluate AI solutions, improving patient outcomes through data-driven strategies.
* Applied advanced analytics and machine learning techniques to AIOps solutions, optimizing infrastructure management and predictive insights for large-scale enterprise environments.
* Automated manual data workflows and reporting processes using **Python** and SQL, enhancing operational efficiency and accuracy.
* Designed and managed data pipelines utilizing **SQL Server**, **Teradata**, and **Hadoop** for clinical and customer data integration.
* Built and optimized web-based applications using **HTML**, **CSS**, **JavaScript**, and **Perl** for data visualization and reporting.
* Integrated web analytics tools such as **Google Analytics** and **Adobe Analytics** for digital performance evaluation.
* Partnered with Fraud and Risk teams to develop anomaly detection models, identifying and mitigating threats to healthcare assets.
* Streamlined **build and release pipelines** for AI solutions using **Kubernetes on OpenShift**, Jenkins, and GitLab CI/CD.
* Synthesized data into strategic insights and presented recommendations to senior leadership through compelling storytelling and visualizations.
* Developed compelling, data-driven presentations and reports using PowerPoint, Figma, and Canva to communicate technical findings and strategic recommendations to stakeholders.
* Developed self-service dashboards and automated metrics for internal stakeholders, increasing accessibility and decision-making efficiency.
* Leveraged Azure Databricks for scalable data processing and model training, integrating seamlessly with Azure Machine Learning for end-to-end machine learning workflows.
* Collaborated with Azure data engineers and architects to optimize data pipelines and infrastructure for AI/ML workloads on the Azure platform.
* Processed annotated data into structured datasets suitable for model training and evaluation, leveraging domain-specific knowledge to enhance data quality and relevance.
* Implemented and validated natural language processing (NLP) models tailored to the clinical domain, utilizing techniques such as language modeling, word embeddings, and text categorization to extract meaningful insights from unstructured medical text data.
* Demonstrated proficiency in applying a diverse range of machine learning algorithms, including Logistic Regression, Decision Trees, SVM, Naive Bayes, kNN, Random Forests, and Dimensionality Reduction Algorithms, to solve clinical NLP tasks effectively.
* Constructed training, testing, and validation datasets for model development, ensuring robustness and generalization capabilities through rigorous evaluation methodologies.
* Developed production-ready model code in Python, adhering to best coding practices and incorporating modular design principles to facilitate scalability and maintainability.
* Utilizing data visualization tools such as Tableau, PowerBI, and matplotlib to create intuitive and informative dashboards and visualizations, effectively conveying complex data insights to non-technical stakeholders.
* Maintained and iteratively improved model codebase, incorporating feedback from performance evaluations and real-world deployment scenarios to enhance model efficacy and reliability.
* Provided mentorship and guidance to junior team members, sharing insights and best practices in NLP modeling techniques, algorithm selection, and model development processes.
* Possess a graduate degree (or equivalent experience) in Computer Science, Machine Learning, or a related quantitative field, demonstrating a strong foundation in advanced mathematical and statistical concepts.
* Applied various AI/machine learning algorithms and statistical modeling techniques such as decision trees, text analytics, and supervised/unsupervised learning to analyze and interpret data.
* Played a key role in the development of Generative AI Platform Capabilities, contributing to the advancement of cutting-edge technologies in artificial intelligence.
* Took charge of delivering AI models to both on-premises infrastructure and cloud platforms, including GCP-Vertex AI and Azure ML, ensuring seamless deployment and optimal performance across diverse environments.
* Used Principal Component Analysis (PCA) for feature engineering and data analysis of high-dimensional datasets, ensuring effective model performance.
* Conducted data cleaning, feature scaling, and engineering using pandas and numpy packages in Python, and built predictive models using deep learning frameworks such as TensorFlow and Keras.
* Deployed machine learning and statistical models on cloud platforms like Amazon Sage Maker and Amazon Forecast, integrating them into existing health insurance systems for real-time decision-making and customer support.
* Developed advanced demand forecasting models for optimizing inventory and pricing strategies, utilizing boosting algorithms like XGBoost and AdaBoost to enhance model performance and accuracy.
* Implemented various machine learning algorithms and statistical modeling techniques including Decision Tree, Text Analytics, Sentiment Analysis, Naive Bayes, Logistic Regression, and Linear Regression, using Python for accuracy rate determination.

**Sr. Data Scientist || Equifax, Alpharetta, Georgia || Jan 2021 – Aug 2022**

# Roles & Responsibilities:

* Designed and implemented **interactive dashboards** using **Tableau** and **Tableau Server**, delivering actionable insights on customer behavior and credit risk assessment.
* Leveraged **SQL**, **Python**, and **R** to extract, analyze, and process large datasets from systems such as **Teradata**, **Oracle**, and **SQL Server**.
* Applied **machine learning algorithms** for predictive modeling and anomaly detection, optimizing customer targeting and reducing fraud risk.
* Conducted **A/B testing** using tools like **Optimizely** and **Adobe Target**, establishing sample sizes and measuring statistical significance for customer experience enhancements.
* Automated manual workflows and developed scalable data pipelines using **Python**, **Hadoop**, and SQL, improving efficiency in data operations.
* Collaborated with marketing, product, and IT teams to define KPIs and build self-service dashboards for cross-functional insights.
* Integrated web analytics tools such as **Google Analytics** and **Adobe Analytics** for performance tracking and customer engagement analysis.
* Deployed data pipelines and machine learning models in hybrid environments using **Kubernetes on OpenShift** and **Docker**.
* Built web-based applications utilizing **HTML**, **CSS**, **JavaScript**, and **Perl** to support analytics and reporting needs.
* Partnered with risk and compliance teams to develop anomaly detection models, mitigating threats and enhancing fraud detection systems.
* Streamlined **build and release pipelines** for analytics tools and AI applications, leveraging **Jenkins** and **GitLab CI/CD** for deployment.
* Delivered insights and recommendations to senior leadership through data storytelling and compelling visualizations.
* Conducted data analysis using Hive for Hadoop cluster data retrieval, SQL for Oracle database data retrieval, and ETL for data transformation.
* Leveraged MLlib, Spark's Machine Learning library, to construct and evaluate various models.
* Devised rule-based expertise systems derived from exploratory analysis and insights from cross-departmental collaboration.
* Conducted Data Cleaning, feature scaling, and feature engineering using pandas and numpy packages in Python, alongside model development using SAP Predictive Analytics.
* Developed MapReduce pipelines for feature extraction using Hive and Pig.
* Created Data Quality Scripts in SQL and Hive for validating successful data loads and data quality, supplemented with diverse data visualizations using Python, Tableau, or Spotfire.

# Data Scientist || Fractal, New York City || Oct 2019 – Dec 2020

# Roles & Responsibilities:

* Developed marketing-focused knowledge graphs to uncover customer journey patterns, optimizing recommendation strategies and enhancing campaign targeting effectiveness.
* Designed scalable machine learning pipelines using Apache Spark, enabling efficient model training and data processing across retail marketing data sources.
* Created Power BI dashboards to communicate data insights to stakeholders, simplifying complex data points into actionable insights for marketing teams.
* Conducted data mining and segmentation analyses to improve audience targeting, significantly enhancing marketing ROI through data-driven decisions.
* Collect and integrate data from various sources including IoT sensors, equipment telemetry, maintenance logs, and historical maintenance records.
* Provided Configuration Management and Build support for over 5 applications, overseeing deployment to production and lower environments.
* Implemented public segmentation through unsupervised machine learning algorithms, employing k-means algorithms with PySpark.
* Utilized R and Python for Exploratory Data Analysis, A/B testing, and hypothesis testing to evaluate the effectiveness of Creative Campaigns.
* Conducted group campaigns using Text Analytics, creating clusters to control and test groups.
* Employed Sentiment Analysis to create positive and negative clusters from merchant transactions, ensuring transaction authenticity and resolving chargebacks.
* Utilized Airflow to monitor job statuses in repositories like MySQL and PostgreSQL databases.
* Extracted data from source XML in HDFS, employing ETL techniques for data preparation and exploratory analysis through data munging.
* Managed Data mapping activities from Source systems to Teradata, including text mining and model building for topic analysis and sentiment analysis on semi-structured and unstructured data.
* Managed data importation from various sources, conducted transformations using Hive, MapReduce, and loaded data into HDFS.
* Developed classes and web page elements using C#, AJAX, and JSP for client-side response validation and database connectivity.
* Created LINUX Shell scripts using NZSQL/NZLOAD utilities to load data from flat files to Netezza database.
* Developed triggers, stored procedures, functions, and packages using Pl/SQL for efficient data processing and manipulation.
* Generated various data visualizations using R, Python, and Tableau/Spotfire, integrating Pipeline Pilot with Spotfire for interactive business-driven layouts.
* Utilized Python, R, and SQL to build Statistical algorithms, including Multivariate Regression, Linear Regression, Logistic Regression, PCA, Randomforest models, Decision trees, and Support Vector Machine for welfare dependency risk estimation.
* Identified and targeted high-risk welfare groups using Machine learning/deep learning algorithms.
* Conducted analyses such as regression analysis, logistic regression, discriminant analysis, and cluster analysis using SAS programming.

# Data Scientist|| Fannie Mae, Virginia || Mar 2018 – Sep 2019

# Roles & Responsibilities:

* Led ETL processes to aggregate and transform marketing and sales data, supporting in-depth analysis and reporting in SQL Server.
* Utilized Tableau for visualization, creating interactive dashboards that allowed marketing and business teams to track key performance metrics in real time.
* Conducted A/B testing and regression analyses to evaluate marketing strategies, identifying opportunities for process adjustments and enhanced marketing effectiveness.
* Integrated multiple data sources to support analytics-ready datasets, improving marketing insights and aligning them with strategic goals.
* Conducted data cleaning and feature selection leveraging the MLlib package in PySpark to ensure data accuracy and relevance for analysis.
* Applied various machine learning algorithms and statistical modeling techniques such as Decision Trees, Text Analytics, Sentiment Analysis, Naive Bayes, Logistic Regression, and Linear Regression using Python to assess model accuracy rates.
* Evaluated and selected the most accurate prediction model based on performance metrics to enhance decision-making.
* Utilized text-mining techniques on customer reviews to identify customer concentrations and extract valuable insights for business improvement.
* Provided analytical support for hotel recommendations and conducted online A/B testing to assess the impact of different strategies on user behavior and preferences.
* Designed Tableau visualizations including bar graphs, scatter plots, and geographical maps to generate detailed summary reports and dashboards for comprehensive data analysis and presentation.
* Developed a hybrid model aimed at enhancing prediction accuracy rates, leveraging insights from various machine learning algorithms and statistical techniques.

# Python Developer/ Programmer Analyst || Koi Reader, Bengaluru, India || Jun 2016 -Jan 2018

# Roles & Responsibilities:

* Developed comprehensive frontend and backend modules utilizing Python within the Django Web Framework, ensuring efficient and scalable application architecture.
* Implemented the presentation layer with proficiency in HTML, CSS, and JavaScript, ensuring a seamless user experience and responsive design.
* Authored stored procedures using Oracle to enhance database functionality and optimized database queries to improve performance and efficiency.
* Designed and implemented a robust data management system utilizing Oracle, ensuring secure and efficient data handling and storage.
* Facilitated effective communication with stakeholders to gather project requirements, ensuring alignment with project goals and objectives.
* Utilized MySQL DB package and Python-MySQL connector for executing numerous MYSQL database queries seamlessly from Python, enabling smooth data retrieval and manipulation.
* Engineered Client/Server applications using C#, JSP, and SQL, delivering reliable and scalable solutions for end-users.
* Created and optimized functions, triggers, views, and stored procedures using MySQL, enhancing database performance and maintenance efficiency.
* Collaborated closely with backend developers to explore and implement innovative solutions within existing Web technology, fostering a collaborative and innovative development environment.
* Participated actively in code review meetings to ensure code quality and adherence to best practices, contributing to the overall improvement of development processes and standards.

**EDUCATION:**

**Dr. A.P.J. Abdul Kalam Technical University Uttar Pradesh, India.**

***Bachelor of Technology* May 2016**