**A logo of a cloud engineer

Description automatically generatedVardhan**

**Sr. Data Scientist**

**Email: raovardhan59@gmail.com**

**Ph:** **+1 (312) 248 3072**

**PROFESSIONAL SUMMARY:**

* Around 8+ years in Development and implemented machine learning algorithms and statistical models to drive product development and improvement, leading to a significant increase in efficiency and effectiveness.
* Created experimental and analytic plans for data modeling processes, enhancing predictive accuracy, and supporting informed business decision-making.
* Leveraged Python coding skills to build robust data structures and algorithms, improving the speed and quality of data processing.
* Utilized SQL and the Hadoop stack for efficient data management and analysis, ensuring timely access to critical data for business stakeholders.
* Conducted advanced analysis with a focus on classification, regression, and clustering methods, providing insightful solutions to complex business problems.
* Explored applications of NLP and deep learning methods in business contexts, contributing to the development of innovative AI-driven tools and processes.
* Python (proficient) with major ML libraries, R, SQL, MATLAB, Java, Scala, C, and C++. Expertise in Python coding for building robust data structures and algorithms.
* Deep experience with opensource Gen AI LLMs e.g., Llama and Dolly. Proficient in leveraging Generative AI and Azure Open AI to drive innovation and efficiency in data science projects. Demonstrated proficiency in developing models using machine learning, deep learning, and advanced techniques like GPT and LLM for tasks.
* Strong understanding of supervised and unsupervised learning, neural networks, transformers.
* Hands-on experience with TensorFlow, PyTorch, SpaCy for rapid prototyping and production-level deployments.
* Expertise in prompt engineering for LLMs, crafting effective prompts to generate desired outputs and improve model interactions.
* Proficient in creating dashboards and visual analytics using Tableau, Python (Matplotlib, Seaborn), and R (ggplot2).
* AWS, Azure, GCP experience. Proficient in Google Cloud Platform (GCP) showcasing comprehensive knowledge and hands-on experience in BigQuery, Vertex AI, Anomaly Detection, and AI/ML integrations.
* Collaborated with cross-functional teams to integrate machine learning solutions into business operations, promoting data-driven decision-making across the organization.
* Exceled in providing data analytics support and improvements, coordinating in research activities for supporting the top management and functional teams with crucial information backup.
* End to end project planning and execution experience for delivering optimal value and implementing best technological solutions to streamline operational procedures.
* Translated client requirements into technical requirements and recommended complete IT solutions.
* Expertise in developing predictive analysis and descriptive analytics as per Specifications.
* Highly methodical in approach and possess quick grasping abilities with keen interest to upgrade skills by acquiring knowledge.
* Understood, implemented, managed, and maintained analytical solutions and techniques independently.
* Well versed with Linear/non-linear, regression and classification modeling predictive algorithms.
* Assessed the cause-effect relationships in data to derive valuable insights and make accurate predictions, providing a solid foundation for strategic planning.
* Participated in the continuous improvement of data governance, contributing to the development of data cleansing and manipulation practices, and ensuring the quality and integrity of data used in analytics.
* Created dashboards as part of Data Visualization using Tableau.
* Experienced in software development life cycle (SDLC) in Agile and Scrum methodologies.
* Expertise and Vast knowledge of Enterprise Data Warehousing including Data Modeling, Data Architecture, Data Integration (ETL/ELT) and Business Intelligence.
* Developed Logical Data Architecture with adherence to Enterprise Architecture.
* Team player with good logical reasoning ability, coordination, and interpersonal skills.
* Proficient in leveraging Generative AI and Azure Open AI to drive innovation and efficiency in data science projects..
* Designed and implemented effective database solutions and models.
* Provided solutions for churn reduction and customer retention.
* Specialized in churn analysis, customer profiling, and interpreting various data points such as billing, loyalty metrics, contact center interactions, and product usage to drive business strategies.
* Worked closely with a major telecommunications client to understand their specific operational challenges and data analytics needs.
* Conducted deep-dive analyses of network utilization and performance metrics to identify bottlenecks, resulting improvement in network efficiency.
* Analyzed customer behavior and usage patterns to identify factors contributing to customer churn. Implemented targeted retention strategies that resulted in a significant reduction in churn rates.
* Implemented advanced analytics and anomaly detection techniques to identify and mitigate fraudulent activities such as SIM card cloning and subscription fraud, safeguarding revenue streams.
* Developed machine learning models for dynamic pricing and personalized service recommendations, resulting in increased average revenue per user (ARPU) and improved customer satisfaction.
* Utilized data analytics to identify upsell and cross-sell opportunities, contributing to increase in average revenue per user (ARPU)
* Deep expertise in developing models using machine learning, deep learning, and advanced techniques like GPT and LLM for tasks.
* Deep proficiency in Google Cloud Platform (GCP), showcasing comprehensive knowledge and hands-on experience in BigQuery, Vertex AI, Anomaly Detection, and AI/ML integrations.
* Acquired knowledge in AWS, highlighting the versatility in cloud platform familiarity.
* Possess foundational knowledge in database management and administration, ensuring data integrity and availability in GCP environments.
* Skilled in integrating LLM platforms into existing frameworks, ensuring seamless functionality and enhanced performance.
* Proficient in a range of neural network architectures including BERT family of Transformer models, BiLSTMs, and RNNs, tailored to suit various NLP tasks.
* Equipped with techniques to handle unbalanced datasets through resampling, weighted loss functions, and other advanced techniques.
* Hands-on experience with industry-leading frameworks such as TensorFlow, PyTorch, and SpaCy for rapid prototyping and production-level deployments.
* Comprehensive experience in the full data science lifecycle including designing experiments, data collection, model training, validation, deployment, and performance management.
* Demonstrated experience in developing and optimizing vectorization pipelines for data transformation and analysis.
* Proficient in enhancing model performance through explicit instruction-based techniques.
* Experienced in integrating human feedback into reinforcement learning to refine and optimize model behavior.
* Proficiency in deploying methods like p-tuning, adaptors, and Layer-wise Relevance Analysis (LoRA) for optimal model adaptation without extensive parameter alteration.
* Proficient in using complex formulas, pivot tables, charts, and data visualization tools to derive actionable business insights from large datasets.

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| Statistical Methods | Hypothesis Testing, ANOVA, Principal Component Analysis (PCA), Correlation (Chi-square test, covariance), Pyomo, Multivariate Analysis, Bayes Law, Time Series Analysis (ARIMA, SARIMA). |
| Machine Learning | Linear Regression, Logistic Regression, Naive Bayes, Decision Trees, Random Forest, Support Vector Machines (SVM), K-Means Clustering, K-Nearest Neighbors (KNN), Random Forest, Gradient Boosting Trees, Ada Boosting, PCA, LDA, Sentiment Analysis, K-Means Clustering, Natural Language Processing, Time Series Forecasting. |
| Deep Learning | Artificial Neural Networks, RBM, DBN, Convolutional Neural Networks, RNN, Deep Learning on AWS, Keras API. |
| Data Visualization | Tableau, Python (Matplotlib, Seaborn), R(ggplot2),QlikView |
| Marketing Analytics | Churn Analysis, Customer Profiling, Billing, Loyalty, Product Usage Analysis |
| Languages | Python (NumPy, Pandas, Scikit-learn, Matplotlib, Seaborn), R, SQL, MATLAB, Java, Scala, C and C++. |
| Operating Systems | UNIX Shell Scripting (via PuTTY client), Linux, Windows, Mac OS |
| Other tools and technologies | TensorFlow, Keras, Pytorch, Hugging Face, Dataiku, AWS ML, NLTK, SpaCy, MS Office Suite, GitHub, AWS, Docker, Sage Maker, BERT, GPT, LoRA |
| Clouds | AWS, GCP, Azure |

**WORK EXPERIENCE**

**Client** **Walgreens, IL March 2020– Present**

**Role: Sr. Data Scientist**

**Responsibilities:**

* Understanding the business problem, framing problem statement, and inferencing insights out of the given data
* Performed feature engineering and handled imbalanced data.
* Extracted hidden value insights and enriched accuracy of the data sets using various statistical models.
* Leveraged technology and automated workflows creating modernized operational processes aligning with the team strategy.
* Spearheaded the integration and fine-tuning of Retrieval-Augmented Generation models to significantly enhance the question-answering capabilities of internal AI systems.
* Utilized RAG to dynamically retrieve and leverage relevant external knowledge, improving the accuracy and depth of responses provided by the system.
* Led the development of cutting-edge NLP applications, incorporating RAG models to generate contextually enriched and informative content.
* Demonstrated the ability to merge external knowledge bases with generative models to create high-quality, relevant outputs tailored to specific user inquiries.
* Directed cross-functional teams in the application of advanced AI and machine learning technologies, including RAG and transformer models, to address complex business challenges.
* Ensured the successful deployment of scalable, impactful solutions that leveraged external knowledge sources for enhanced decision-making and customer engagement.
* Used TensorFlow and keras libraries for preprocessing and model building.
* Applied different classification models like logistic regression, SVM and Ensemble models – Random Forest and XGBoost.
* Performed Data cleaning and exploratory data analysis on the data & developed insights which was refined by collaborating with the SMEs.
* Led the integration and fine-tuning of Generative AI solutions, including Azure Open AI, enhancing NLP applications and system capabilities.
* Designed, developed, and implemented advanced AI solutions to solve complex business problems across various domains.
* Collaborate with engineers, researchers, and product managers to understand business needs and translate them into technical specifications for AI models.
* Gathered, cleaned, and prepared data for training and evaluating sophisticated AI models, ensuring data quality and ethical considerations.
* Explored and implemented state-of-the-art AI techniques, including deep learning, natural language processing, and computer vision, tailored to specific needs.
* Evaluated and compared different AI model architectures and hyperparameters to optimize performance, address potential biases, and ensure responsible development.
* Developed monitoring and evaluation frameworks to track AI model performance and ensure alignment with business goals.
* Communicated complex AI concepts and findings to technical and non-technical audiences through data visualization, storytelling, and presentations.
* Spearheaded the deployment of ML models at scale, focusing on operational efficiency and robust data science solutions.
* Evaluated and fine-tuned Large Language Models (LLM) for specific industry applications, optimizing for accuracy and performance.
* Led the integration and fine-tuning of open-source Generative AI LLMs, specifically Llama and Dolly, to enhance the natural language processing capabilities of our AI systems.
* Utilized Llama and Dolly for developing advanced AI solutions, including the creation of contextually enriched and informative content.
* Directed cross-functional teams in the application of Llama and Dolly, alongside other advanced AI and machine learning technologies, to address complex business challenges.
* Developed and deployed healthcare-focused ML models, contributing to predictive analytics and patient care optimization.
* Performed feature scaling, feature engineering and statistical modeling using various python packages like scikit-learn, SciPy, feature tools etc.
* Built forecasting models to capture patterns from time series data by using traditional forecasting models like ARIMA, Holt-Winters and machine learning techniques like XGBoost, LightGBM etc. using azure machine learning studio.
* Collaborated closely with business stakeholders to thoroughly understand complex business problems, framed precise problem statements, and extracted meaningful insights from available data.
* Implemented feature engineering techniques to enhance model performance and effectively managed imbalanced datasets to improve model accuracy.
* Spearheaded the fine-tuning process of large language models, ensuring optimized performance for bespoke tasks and specific industry domains.
* Applied a diverse range of classification models, including logistic regression, Support Vector Machines (SVM), and ensemble models such as Random Forest and XGBoost, to solve complex classification problems.
* Demonstrated proficiency in leveraging technology to automate workflows, resulting in the modernization of operational processes in alignment with team strategy.
* Successfully deployed end-to-end pipelines using Azure Data Factory, with triggers for timely execution of ML studio experiments and storage of forecasted results in SQL tables.
* Collaborated with sales and marketing teams to align pricing strategies with market conditions.
* Implemented A/B testing to evaluate pricing models' effectiveness.
* Skilled in prompt engineering, fine-tuning, and deploying generative models to enhance AI systems' capabilities.
* Integrated external factors like economic indicators and market trends into forecasting models.
* Applied machine learning techniques for enhanced forecast accuracy.
* Developed custom dashboards for real-time demand tracking and decision-making.
* Worked closely with supply chain teams to align production with forecasted demand.
* Deployed an end-to-end pipeline using Azure Data Factory and created trigger for the timely running of the ML studio experiment flow and stored the forecasted results in SQL table.
* Worked closely with various stakeholders to collect, clean, model and visualize datasets.
* Develop Q&A Models and Worked on Building Voice Synthesis Model for Choice Internal Bot.
* Leveraged technology and automated workflows creating modernized operational processes aligning with the team strategy.
* Competent in using JIRA for project management, ensuring smooth workflow and effective team collaboration.
* Skill in tuning hyperparameters specific to Merlin models, such as model size, learning rate, and batch size, to achieve optimal performance.
* Used TensorFlow and keras libraries for preprocessing and model building.
* Proven track record of working alongside software engineers to implement and test high-quality code for data analytics and forecasting models.
* Established high performing globally Connected and innovative mind set in the team culture.
* Pioneered the integration of Gen AI in various business processes, driving automation and enhancing decision-making capabilities.
* Experienced in deploying statistical and machine learning models on cloud platform like Amazon SageMaker for enhanced model performance and scalability.
* Developed and optimized machine learning models using Amazon SageMaker, enhancing accuracy and efficiency.
* Automated ML workflows with SageMaker, significantly reducing manual effort and improving model performance.
* Deployed and managed scalable ML models in production environments using SageMaker's deployment capabilities.
* Integrated SageMaker with other AWS services (like S3, Lambda) for comprehensive data solutions.
* Led the design and implementation of generative AI solutions, producing state-of-the-art models tailored for diverse applications.
* Spearheaded the fine-tuning process of large language models to tailor them for specific industry domains, significantly enhancing performance and application relevance.
* Implemented LangChain to seamlessly integrate advanced language processing capabilities into business operations, driving innovation and efficiency.
* Conducted advanced research on generative algorithms, staying abreast of industry advancements and ensuring best-in-class AI utilization.
* Established a robust framework for LLM fine-tuning, ensuring consistent performance gains across varied tasks.
* Collaborated with ethics and regulatory teams to ensure LLMs align with industry standards and organizational values during the fine-tuning process.
* Developed custom tokenization and embedding strategies, tailored to specific domain lexicons and jargons.
* Designed rigorous evaluation metrics and benchmarks to assess the performance of fine-tuned models against standard models.
* Worked closely with stakeholders to capture domain-specific knowledge, ensuring LLMs resonate with target audience nuances.

**Environment:** Machine Learning, AWS, GCP, Python (Scikit-learn, SciPy, NumPy, Pandas, Matplotlib, Seaborn), GitHub, Linux, Tableau.

**Client: Asurion, TN Jan 2019 – March 2020**

**Role: Data Scientist**

**Responsibilities:**

* Performed Data cleaning and exploratory data analysis on the data using Apache Spark SQL and crunched the data by creating data transformations to be used for further model building.
* Analyzed data using SQL, Python, R, Apache Spark and presented analytical reports to technical teams.
* Used Scikit-Learn’s model selection framework to perform hyper-parameter tuning using GridSearchCV and RandomizedSearchCV algorithms.
* Designed and implemented robust ontology frameworks to structure and standardize complex data sets.
* Created semantic models that define the relationships and classifications within data, enhancing data consistency and interoperability across various systems and platforms.
* Experienced in prompt engineering for LLMs, crafting effective prompts to generate desired outputs and improve model interactions.
* Developed advanced AI models for pricing and forecasting, significantly improving decision-making processes and operational efficiency.
* Demonstrated excellence in growing and managing onshore contract resources, ensuring high performance and operational efficiency.
* Fostered a culture of continuous learning and development, enabling teams to achieve ambitious project goals.
* Led pricing/forecasting projects, delivering actionable insights and automated workflows that streamlined operational processes.
* Contributed to dynamic pricing strategies, customer retention efforts, and supply chain optimizations demonstrating the versatility and application of AI/ML models in improving business outcomes.
* Demonstrated capability in deploying and managing scalable machine learning models in cloud environments, including Azure and AWS, for high-traffic applications.
* Proficient in building and evaluating ML models for healthcare, utilizing data analytics to improve diagnostic tools and patient outcomes.
* Designed and executed complex data queries using Hive, enabling insightful analysis for business decision.
* Integrated Hive with other data processing tools for streamlined data manipulation and reporting.
* Developed and deployed LLMs for IRQA systems, allowing them to understand and respond to user queries effectively, ensuring relevant, accurate, and contextual information retrieval and responses.
* Performed hyper parameter tuning by formulating Distributed Cross Validation in Apache Spark to speed up the computation process.
* Clustered time series of similar SKUs using KMeans on extracting features from the time series like trend, mean, standard deviation etc.
* Spearheaded the fine-tuning process of large language models, ensuring optimized performance for bespoke tasks and specific industry domains.
* Employed a heterogeneous stacked ensemble of methods for the final forecast.
* Utilize distributed computing tools like Hadoop, Hive, or Spark.
* Developed and deployed LLMs for a variety of applications, from IRQA systems to customer insights, showcasing versatility in applying AI for business solutions.
* Integrated MLOps principles with a focus on LLM and LangChain technologies, enhancing model performance, scalability, and governance.
* Utilized Git for version control on GitHub and MLFlow for model versioning to collaborate work with the team members.
* Created a data pipeline using Docker and Airflow for running the model every week.
* Monitored model performance by building a dashboard in Dash.
* Researched, experimented, and built time series models to better capture patterns for intermittent time series and cluster similar time series using K-Means on global features like wavelet transforms and KShape clustering.
* Built statistical models like constrained regression to capture price elasticity.
* Collaborated with business stakeholders for capturing product requirements, sketch product roadmaps, understand market specific nuances and help adopt the product in 5 countries replacing a third-party tool.
* Improved forecasting performance by 27% (sMAPE) on average across countries.
* Reduced computational time by 70% (10 hours to 3 hours) by using distributed computation i.e., UDFs on Spark and RAPIDS on GPUs for building 10 million statistical/ machine learning models per country using the same computational resources.
* Integrated MLOps principles utilizing Data/Model Governance, Code/Data/Model Versioning and built Data Pipelines using open-source tools and was awarded the SPOT Award for the same.
* Built a customer insights product for capturing Walmart's store concerns like COVID mask concerns, parking space, long queues etc. using tweets from Twitter utilizing an open-source package Twint. Also, built a Tableau dashboard for the customer team to provide insights of major concerns in stores in USA.
* Developed and deployed multiple NLP models using machine learning and deep learning algorithms such as GPT-3 and BERT.
* Implemented real-time analytics to monitor model performance, enabling dynamic tuning and optimization.

**Environment:** Python, Tableau, Machine Learning (Logistic regression/ Random Forests/ KNN/ K-Means Clustering/ Hierarchical Clustering/ Ensemble methods/ Collaborative filtering) GitHub

**Client: Phillips 66, India Jun 2016 – Aug 2018**

**Role: Data Scientist**

**Responsibilities:**

* Performed Data cleaning and exploratory data analysis on the data using Apache Spark SQL and crunched the data by creating data transformations to be used for further model building.
* Analyzed data using SQL, Python, R, Apache Spark and presented analytical reports to technical teams.
* Built forecasting models for forecasting weekly demand of a SKU by capturing patterns from intermittent time series data by using traditional forecasting models like Crostons method, Moving Averages, and machine learning techniques like XGBoost, LightGBM etc. using python.
* Used Scikit-Learn’s model selection framework to perform hyper-parameter tuning using GridSearchCV and RandomizedSearchCV algorithms.
* Performed hyper parameter tuning by formulating Distributed Cross Validation in Apache Spark to speed up the computation process.
* Clustered time series of similar SKUs using KMeans on extracting features from the time series like trend, mean, standard deviation etc.
* Employed a heterogeneous stacked ensemble of methods for the final forecast.
* Evaluated the performance of our model using accuracy metrics like MAPE, MAD etc.
* Utilized Git for version control on GitHub and MLFlow for model versioning to collaborate work with the team members.
* Created a data pipeline using Docker and Airflow for running the model every week.
* Skillfully combined various algorithms and modules, including Vertex AI, to improve data analysis accuracy.
* Built statistical models like constrained regression to capture price elasticity.
* Collaborated with business stakeholders for capturing product requirements, sketch product roadmaps, understand market specific nuances and help adopt the product in 5 countries replacing a third-party tool.
* Improved forecasting performance by 27% (sMAPE) on average across countries.
* Reduced computational time by 70% (10 hours to 3 hours) by using distributed computation i.e., UDFs on Spark and RAPIDS on GPUs for building 10 million statistical/ machine learning models per country using the same computational resources.
* Integrated MLOps principles utilizing Data/Model Governance, Code/Data/Model Versioning and built Data Pipelines using open-source tools and was awarded the SPOT Award for the same.
* Built a customer insights product for capturing Walmart's store concerns like COVID mask concerns, parking space, long queues etc. using tweets from Twitter utilizing an open-source package Twint. Also, built a Tableau dashboard for the customer team to provide insights of major concerns in stores in USA.

**Environment:** Python, Tableau, Machine Learning (Logistic regression/ Random Forests/ KNN/ K-Means Clustering/ Hierarchical Clustering/ Ensemble methods/ Collaborative filtering) GitHub