Hasan Fa

raviteja@iflexpro.com |**(330) 403-5265**

# Summary

With a proliﬁc 10+ years of experience as a profound Data Scientist and Machine Learning Engineer, I bring unparalleled expertise in advanced predictive modeling, deep learning, Artiﬁcial Intelligence(AI), and statistical programming. With outstanding proﬁciency in Python, R, SQL, Tableau, Power BI, and many data science, machine learning, and business tools, I bring extensive experience while collaborating with cross-functional teams to solve multifaceted business challenges. This includes building scalable ML & AI pipelines, complex problem-solving, robust feature engineering, dynamic time-series forecasting, and impactful dashboard creations, mentoring junior data scientists, and evangelizing a data-driven culture. With my analytical rigor and leadership abilities, I am poised to accelerate value creation as a Lead Machine Learning Engineer.

# Experience

## Machine Learning Engineer Sep 2023 – Present

*Naturade (Irvine, CA) (Remote)*

Programming and Other Tools Used: Python (NumPy, Pandas, Seaborn, Scikit-learn(sklearn), Google Analytics, SpaCy, openCV, matplotlib, Anaconda AI/ML libraries), SQL, TensorFlow, Keras, Docker, Kubernetes, Jenkins pipeline & other DevOps tools, Git Version Control, Pytorch, AWS platforms, Tableau, Power BI, Power Automate, ETL pipelines, Databricks, Stripe Payments

* Engineered robust machine learning models for product recommendation systems, enhancing personalization accuracy.
* Built ETL pipelines for automated data ingestion, preprocessing, and feature extraction.
* Conducted in-depth data wrangling and statistical analysis for insight generation and anomaly detection.
* Utilized TensorFlow, PyTorch, and Scikit-Learn for model training, tuning, and evaluation.
* Implemented real-time prediction APIs using Flask and FastAPI for seamless model integration.
* Optimized model performance through hyperparameter tuning, regularization, and feature engineering.
* Leveraged cloud platforms (AWS, GCP) for scalable model deployment and pipeline automation.
* Monitored model drift and retrained pipelines using MLOps frameworks (MLflow, Kubeflow).
* Developed scalable architectures for high-volume data processing with Spark and Hadoop.
* Documented ML workflows, scripts, and experiments for reproducibility and team collaboration.

## Lead Data Scientist/ML Engineer Oct 2019 – Sep 2023

*Shrewd Food (New York, NY)*

Programming and Other Tools Used: Python (NumPy, Pandas, Seaborn, Scikit-learn(sklearn), Google Analytics, SpaCy, openCV, matplotlib, Anaconda AI/ML libraries), R(Shiny, cluster, dplyr, tidyr, infer, recommenderlab), SQL, TensorFlow, Keras, Docker, Kubernetes, Jenkins pipeline & other DevOps tools, Git Version Control, Pytorch, AWS platforms, Tableau, Power BI, Power Automate, ETL pipelines, Databricks, Stripe Payments

* Developed advanced predictive models( ranging from Exponential Smoothing to building neural networks & using Support Vector Machines, Random Forest Algorithms), built Machine Learning models to forecast subscription trends, churn prediction, audience segmentation, optimizing inventory levels and price optimization.
* Headed the data science team, orchestrating cross-functional efforts to leverage data for advanced analytics and informed decision-making. Managed the policies & roles for the team with IAM to give different access levels.
* Conducted comprehensive customer proﬁling, leveraging advanced segmentation techniques, predictive modeling, and data enrichment, encompassing customer scoring, prospect analysis, lead generation, data mining, market and trade area analysis, distribution channel assessment, and performance tracking, feeding into streamlit, culminating in rigorous ROI analysis using AWS EMR, Athena and Amazon QuickSight.
* Leveraged cohort analysis utilizing Python and SQL to dissect subscriber behavior patterns on Shrewd Food DTC channel, enhancing personalized recommendations and upsell opportunities, while validating outcomes through A/B testing and continuous reﬁnement.
* Collaborated with the Marketing team to design and analyze A/B tests, enhancing subscription conversion rates through data-driven pricing and promotional strategies.
* Deeply worked on Data sourcing, Data Wrangling, Data Ingestion with ETL pipelines, and manipulation across diverse sources, ensuring precise and streamlined data integration for robust reporting and analysis pipelines.
* Employed state-of-the-art Natural Language Processing (NLP) algorithms, including tokenization, part-of-speech tagging, and sentiment analysis, to meticulously dissect and interpret nuanced customer feedback, driving iterative product enhancements and elevating overall customer satisfaction levels.
* Leveraged the power of Tableau, Google Data Studio, Power Automate and Power BI to architect and develop dynamic and interactive dashboards, meticulously displaying and tracking intricate subscription metrics, thereby facilitating seamless real-time monitoring of critical key performance indicators essential for informed decision-making.
* Collaborated with Product Development to identify gaps and opportunities in the subscription offering, utilizing data insights to drive innovation and product enhancements.
* Employed time series analysis to identify seasonality and demand fluctuations, optimizing production schedules for efﬁcient resource allocation.
* Leveraged Python's extensive libraries, including pandas for data preprocessing, feature tools for comprehensive feature engineering, and sklearn for intricate model development, ensuring the creation of highly resilient and scalable data science solutions.
* Develop churn prediction models using classiﬁcation algorithms to proactively identify and address potential subscription cancellations.
* Collaborate with the Operations team to streamline supply chain processes, minimizing subscription delivery lead times through demand forecasting.
* Implement collaborative ﬁltering techniques to enhance personalized subscription recommendations, driving customer engagement and loyalty.
* Performed digital marketing analytics by analyzing pixel and cookie data and modeled segmentation algorithms using machine learning techniques.
* Developed an AI & ML driven sentiment analysis system using NLP techniques with spaCy and gensim, and implemented a GAN-based art generation model using TensorFlow GAN
* Conduct preliminary proof of concepts, outline initial data wrangling steps, and offer direction for developing the production-ready version of the functionalities.
* Led and guided a high-performing team of data scientists, fostering innovation, driving data-driven decision-making, and achieving impactful business outcomes. This role also involves tasks such as recruiting, coaching, training, and supporting the career growth of team members.
* Contributed to the implementation of a robust DevOps CI/CD pipeline for automated deployment of a real-time data analytics web application using Docker and Kubernetes, ensuring faster releases and improved collaboration.

## Achievement:

* + Developed & validated 8+ models and optimized well over 50 lead-gen & retention funnels and pipelines to catapult growth, resulted in 44% increase in ROAS, an 82% increase in LTV, and a 48% decrease in churn rate for subscriptions.
	+ Led a team of data science, machine learning & AI engineers, Modeled complex machine learning algorithms that segments customers in real time for precision targeting and offers a personalized brand experience, resulting in a 40% increase in Average Revenue Per User(ARPU).

## Data Scientist Aug 2016 – Oct 2019

*The Cigna Group (Bloomﬁeld, CT)*

Programming and Other Tools Used: SQL, Python (Modules: Pandas, Numpy, Matplotlib, Seaborn, SciPy, sklearn), R (Packages: caret, dplyr, ggplot2, forecast, zoo, reshape2, stringr etc), PySpark MLlib, Azure Data Factory, Tableau, Alteryx, PowerPivot, sagemaker, Power View

* + Developed and deployed predictive models using machine learning algorithms (e.g., XGBoost, Random Forest) to forecast patient health outcomes, enabling proactive interventions and personalized treatment plans.
	+ Conducted in-depth analysis of clinical data, applying statistical methods and hypothesis testing to identify trends, patterns, and insights that inform medical decision-making and patient care strategies.
	+ Utilized NLP techniques to extract valuable information from unstructured EHR data, enabling efﬁcient extraction of patient histories, diagnoses, and treatment progress for holistic healthcare assessments.
	+ Develop anomaly detection models and network analysis techniques to detect fraudulent activities within insurance claims data, safeguarding the integrity of the healthcare system.
	+ Managed and analyzed datasets containing PII/PHI, ensuring compliance with data privacy and security regulations.
	+ Employed clustering and segmentation techniques on demographic and health data to categorize patient populations, aiding in the identiﬁcation of at-risk individuals and the development of targeted health promotion initiatives.
	+ Created risk stratiﬁcation models using logistic regression and deep learning architectures to identify individuals at risk of chronic diseases, contributing to early intervention and preventive measures.
	+ Worked closely with medical professionals, data engineers, and IT teams to integrate and preprocess diverse data sources (structured and unstructured) into Azure Data Factory, ensuring seamless data flow for advanced analytics and modeling.
	+ Followed the deﬁned operating procedures to calculate data quality metrics.
	+ Ensure that KBE deﬁnitions and metrics are inventoried, maintained, and improved continuously to follow the change management process to add/ remove/ change the KBE & as needed through the rules using the SDA (service delivery application) data.
	+ Crafted and executed controlled experiments to validate hypotheses and drive data-driven decisions while creating relevant features from raw data to enhance model performance and insights.
	+ Collaborative team player showcasing exceptional communication and adept problem-solving skills. Proﬁcient in expressing analytical concepts and methods through both oral and written communication.
	+ Hands-on involvement in collaborating across diverse business and technical units.

## Data scientist Aug 2015 – Aug 2016

*Techno Script Solutions (Chennai)*

Programming and Other Tools Used: R (packages: caret, dplyr, ggplot2, forecast, zoo, reshape2, stringr, etc), Python (Modules: pandas, Numpy, SciPy, Scikit-Learn), SQL, AWS ecosystem, Tableau, PowerPivot, Power View, SSRS

* + Developed and deployed predictive models using Python and scikit-learn for credit risk assessment, enhancing loan approval accuracy.
	+ Analyzed member transaction data using SQL to uncover spending patterns and recommended personalized ﬁnancial services.
	+ Utilized time series analysis to forecast deposit trends and optimized liquidity management strategies.
	+ Collaborated with marketing teams to design A/B tests using R and caret, optimizing campaign effectiveness and member engagement.
	+ Created interactive Tableau dashboards to track key performance indicators, providing actionable insights for executive decision-making.
	+ Stayed current with emerging data science trends, contributing to innovation and fostering a data-driven culture within the organization.
	+ Create compass point data reports and build price optimization routines

## Quantitative Data Analyst July 2014 – Aug 2015

*Vish Gyana Technology Solutions (Chennai)*

Programming and Other Tools Used: SQL, R (packages: caret, dplyr, ggplot2, forecast, reshape2, stringr, etc), Python, Advanced Excel Ops, Google Cloud Platform, pytorch, Tensorflow

* + Gathering, deﬁning, and documenting requirements, identifying and analyzing metrics and key performance factors for UX tests
	+ Utilized Python(pandas), PowerQuery for data extraction and preprocessing, transforming raw ﬁnancial data into structured datasets for analysis.
	+ Applied statistical methods ( hypothesis testing, regression analysis) to identify and interpret patterns in market trends, supporting investment decisions.
	+ Collaborated with stakeholders to develop pricing models using derivatives valuation techniques (Black-Scholes, Monte Carlo simulations) for accurate risk assessment.
	+ Conducted time series analysis to forecast market volatility, enhancing risk management strategies and informing portfolio adjustments.
	+ Leveraged SQL to query and analyze large-scale transactional data, uncovering anomalies and potential fraudulent activities for fraud detection.
	+ Developed advanced Excel models for portfolio optimization, integrating historical data and risk parameters to recommend balanced investment strategies.
	+ Created Tableau dashboards with interactive visualizations to communicate complex ﬁnancial analyses and insights to stakeholders.
	+ Contributed to the development of algorithms using the data from IoT sensors and ZigBee processors, implementing quantitative models and backtesting using R and MATLAB.
	+ Stayed informed about evolving ﬁnancial regulations, industry trends, and data science advancements to continuously enhance analytical methodologies.

# Technical Skills

## Machine Learning:

Supervised Learning: Linear and Logistic regression, naive Bayes, decision trees, random forest, support vector machines, bagged and boosted decision trees, k-nearest neighbor (k-NN), Neural Networks, Collaborative ﬁltering, gradient boosting machines.

Unsupervised Learning: K-Means, hierarchical clustering, DBSCAN, Principal Component Analysis, t-SNE, spectral clustering, Gaussian Mixture Models, Apriori Algorithm.

**Model Evaluation & Validation:** Stratiﬁed sampling, Mean Squared Error(MSE), ROC curve & AUC, F1-score, Cross Validation Variance, Bias-variance trade-off, imbalanced-learn.

**Other Statistical Skills:** Time-Series Forecasting, Descriptive Statistics, correlations, t-tests, ANOVA, chi square statistical testing, cross-tabs, decision-making theory, Monte-Carlo simulation, text mining, clustering and segmentation, survival & retention analysis.

**Programming:** Python, R, MATLAB, and SQL.

**AI & Generative AI:** SpaCy, NLTK, transformers, IBM Adversarial Robustness Toolbox, OpenAI Gym, TensorFlow-GAN

**Database Tools:** SQL, MS Access, Excel, Airtable, Power Query, Power Pivot, Databricks, PySpark, Airflow, Neo4j, Alteryx, Salesforce Cloud Tools.

**Visualization Tools:** Tableau, ML Flow, Bokeh, Matplotlib, Seaborn, Power BI, R Shiny, Google Analytics, and Google Data Studio, Adobe Experience Cloud.

**Cloud Tools:** Amazon Web Services(AWS) services, Azure (Azure ML products, Azure Data Factory), Google Cloud Platform, serverless computing(Lambda, Cloud functions).