Hasan Fa

raviteja@iflexpro.com | (330)-403-5265 | <https://www.linkedin.com/in/hasanf7/>

**Data Scientist / Machine Learning Engineer**

# Professional Summary

* **8+ years of experience** as a Machine Learning and AI Engineer, specializing in predictive modeling, statistical analysis, and AI solutions.
* Expertise in building and deploying **machine learning models** for **customer engagement, marketing campaigns, and subscription-based systems**.
* Proficient in **Python, SQL**, and data tools like **Dataiku, GCP**, with a strong focus on **A/B testing methodologies**.

## Skilled in developing models for Customer lifetime value (LTV), Conversion Rate optimization, churn prediction, and app personalization.

* Extensive experience in **data integration** and **feature engineering** to create **holistic customer profiles**.
* Expertise in **model deployment** and scaling on cloud platforms, ensuring **performance and scalability** in production environments.
* Collaborative experience with cross-functional teams to deliver impactful insights and drive **data-driven business strategies**.
* Strong leadership abilities, including **mentoring junior engineers** and fostering a

## data-driven culture.

* Adept at working with **complex high-volume datasets**, developing efficient **feature engineering solutions**, and solving complex business challenges.
* Ready to leverage my skills to **enhance user engagement** and **optimize business outcomes** in a data-driven environment.
* Worked closely with **Key Decision makers, Marketing, Sales, Customer Service, and Retails teams.**

# Experience

## Lead Data Scientist / Machine Learning Engineer

*Naturade (BlackRock Investment Portfolio) (Remote)*

* Engineered robust solutions and machine learning models for **e-commerce website and app** experiences, focusing on **product recommendation systems** and **enhancing**

**personalization accuracy** for customers using tools such as **TensorFlow**, **PyTorch**, and

## Scikit-Learn.

* Performed **data wrangling** and **statistical analysis** using **Pandas**, **NumPy**, and **SQL** to generate actionable insights and detect anomalies, which were integrated into the AI system to improve customer interactions.
* Utilized **TensorFlow**, **PyTorch**, and **Scikit-Learn** for **model training, tuning**, and **evaluation**, while implementing **real-time enhancements** using **WebSockets** to optimize chatbot responsiveness.
* Applied **regression analysis** using **scikit-learn** to identify key business drivers and evaluate their impact on various **customer performance metrics**, including **conversion rates**, **churn prediction**, and **subscription renewals**.
* Developed **real-time prediction APIs** with **Flask** and **Google Cloud API Gateway**, ensuring seamless integration of machine learning models into the **website and app** for personalized user experiences.
* Designed and implemented **churn prediction models** using **logistic regression**, **decision trees**, and **random forests** to assess **customer retention**, predict **churn risk**, and optimize **subscription management** strategies.
* Enhanced **subscription renewal** predictions through machine learning by analyzing **user behavior**, **historical interactions**, and **engagement metrics**, improving customer retention and lifetime value.
* Optimized model performance through **hyperparameter tuning**, **regularization**, and **feature engineering** techniques in **Google Cloud AI Platform**, achieving significant improvements in predictive accuracy and personalization.
* Ensured accessibility compliance with **ADA, WCAG**, and other industry standards to make the app and website more inclusive for a diverse customer base.
* Leveraged **Google Cloud Platform (GCP)** tools like **BigQuery**, **Dataflow**, and **AI Platform** for scalable model deployment, pipeline automation, and data storage, allowing efficient processing of customer interactions at scale.
* Implemented **model monitoring** with **Datadog** and tracked **model drift** using **Google Cloud Monitoring**, retraining models with **MLflow** and **Kubeflow** for continuous optimization and performance tracking.
* Developed **scalable data processing architectures** using **Apache Spark** and **Hadoop** on **Google Cloud Dataproc**, enabling the efficient handling of large volumes of customer data, especially in **subscription and churn management**.
* Documented **ML workflows, scripts**, and **experiments** in **Jupyter Notebooks** and version-controlled them using **Git** to ensure reproducibility and enhance team collaboration in an **Agile** development environment.
* Contributed to **A/B testing** frameworks for continuous product improvements, utilizing

**Adobe Analytics** and **Google Optimize** to assess the effectiveness of

**subscription-based features** and **churn mitigation strategies**, directly influencing customer retention initiatives on e-commerce channel.

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, Google Cloud Platform(GCP), Argo CD, Tableau, MLOps, DynamoDB, CI/CD pipelines, SAST tools, Datadog, Adobe Analytics, Stripe Payments

## Lead Data Scientist

*Shrewd Food (New York, NY)*

* Developed advanced **predictive models** using techniques such as **Exponential Smoothing**, **neural networks**, **Support Vector Machines**, and **Random Forest Algorithms** to forecast **subscription trends**, **churn prediction**, **audience segmentation**, **inventory optimization**, and **price optimization**.
* Led the **data science team**, orchestrating cross-functional efforts to leverage

e-commerce data for **advanced analytics** and informed decision-making, managing team policies and roles using **IAM** to define appropriate access levels.

* Conducted comprehensive **customer profiling** through advanced **segmentation techniques**, **predictive modeling**, and **data enrichment**, driving **ROI analysis** via **GCP services** such as **BigQuery**, **Cloud Identity**, **Cloud Storage**, and **Google Data Studio** for performance tracking.
* Leveraged **cohort analysis** with **Python** and **SQL** to analyze **subscriber behavior patterns** on the **Shrewd Food DTC channel**, improving personalized recommendations and upsell opportunities, validated via **A/B testing**.
* Developed a **predictive model** that improved **demand forecasting accuracy** by 25%, incorporating seasonal fluctuations and special events to optimize subscription strategies.
* Applied **SARIMA modeling** for **monthly sales forecasting**, incorporating **seasonal patterns**, **trends**, and **cyclical variations** to drive business decisions.
* Collaborated with the **Marketing team** to design and analyze **A/B tests**, driving subscription conversion rates through data-driven pricing and promotional strategies.
* Led **data sourcing**, **data wrangling**, **data ingestion** with **CI/CD pipelines**, and **data manipulation** across diverse sources to ensure precise data integration and streamlined reporting and analytics.
* Implemented **Natural Language Processing (NLP)** algorithms, including **tokenization**, **part-of-speech tagging**, and **sentiment analysis**, to interpret customer feedback, driving iterative product enhancements.
* Designed and developed **dynamic, interactive dashboards** using **Google Data Studio**, **Power BI**, **Tableau**, and **Power Automate**, facilitating real-time tracking of critical subscription metrics for data-driven decision-making.
* Collaborated with **Product Development** to identify gaps and opportunities in **subscription offerings**, leveraging 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 libraries** such as **pandas** for **data preprocessing**, **featuretools** for **feature engineering**, and **sklearn** for **model development**, ensuring robust and scalable data science solutions.
* Developed **churn prediction models** using **classification algorithms** to proactively identify and address potential subscription cancellations, reducing churn rates.
* Worked closely with the **Operations team** to streamline **subscription delivery lead times** through **demand forecasting**, enhancing supply chain efficiency.
* Implemented **collaborative filtering** techniques for personalized **subscription recommendations**, improving customer engagement and loyalty.
* Performed **digital marketing analytics** by analyzing **pixel** and **cookie data**, using **machine learning** techniques to model **segmentation algorithms** for targeted campaigns.
* Developed an **AI-driven sentiment analysis system** using **spaCy** and **gensim**, and built a **GAN-based art generation model** using **TensorFlow GAN** to enhance product personalization.
* Conducted preliminary **proof of concepts**, outlining initial **data wrangling** steps and providing direction for developing production-ready functionalities.
* Led and mentored a **high-performing team** of **data scientists**, **machine learning engineers**, and **AI engineers**, fostering innovation, driving **data-driven**

**decision-making**, and supporting the career growth of team members.

* Contributed to implementing a robust **MLOps, CI/CD pipeline** for automated deployment of real-time data analytics web applications using **Docker**, **Kubernetes**, and **Google Cloud Build**, improving release times and team collaboration.

**Achievements:**

* Developed and validated **8+ models**, optimizing over **50 lead-gen & retention funnels**, resulting in a **44% increase in ROAS**, **82% increase in LTV**, and **48% decrease in churn rate** for subscriptions.
* Led a team of **data science**, **machine learning**, and **AI engineers**, developing real-time **customer segmentation** algorithms, resulting in a **40% increase in ARPU** by offering personalized brand experiences.

Programming and Other Tools Used: Python, SQL, GCP, Adobe Experience Cloud, Adobe Analytics, Optimizely(A/B Testing), Argo CD, Google Data Studio, Apple App Store & Google Play Store Analytics, Looker Stuio, Google and Meta Ads Platform,

## Data Scientist

*The Cioftna Group (Bloomfield, CT)*

Programming and Other Tools Used: SQL, Python (Pandas, Numpy, Matplotlib, Seaborn, SciPy, sklearn), R (Packages: caret, dplyr, ggplot2, forecast, zoo, reshape2, stringr etc), PySpark MLlib, Tableau, GCP, Vertex AI, BiqQuery, Tensorflow

* 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. Refactored and improved APIs using GCP cloud functions.
* 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. Improved and enforced code quality by refactoring existing code, following best practices, and implementing automated checks.
* 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. Ensured secure coding practices and regulatory compliance. 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

*Techno Script Solutions*

Programming and Other Tools Used: R (packages: caret, dplyr, ggplot2, forecast, zoo, reshape2, stringr, etc), Python (Modules: pandas, Numpy, SciPy, Scikit-Learn), Javascript, Node.js, 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.

# 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, ML Workflows

**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, Javascript (Typescript), SQL, MySQL, MS SQL.

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

**Database:** SQL, Google BigQuery, GCP CLoud Storage, Excel, Airtable, Power Query, Power Pivot, Databricks, PySpark, Airflow, DynamoDB, ACR, 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:** Google Cloud Platform(GCP), Amazon Web Services(AWS) services, Azure (Azure ML products, Azure Data Factory), serverless computing(Lambda, CloudFront, Cognito, CDK, Cloud functions).

# Achievements

## Artic Code Vault Contributor - 2020

A Github’s project to store open-source code for future generations.