**Aparna Hanumantu**

chaitanya@keentechnos.com 469 972 5599

**PROFESSIONAL SUMMARY**

 Experienced data and AI professional with over 10 years of success delivering end-to-end, data-driven solutions across diverse industries. Expertise in building robust ETL pipelines, automating workflows, and developing scalable data architecture using cloud platforms like AWS, Azure, and GCP. Proficient in advanced analytics, machine learning, and AI model development for solving complex business problems and optimizing performance at scale. Strong focus on operational efficiency, including continuous integration, real-time data processing, and intelligent automation. Committed to secure application development with adherence to data governance, compliance, and privacy standards across the full data lifecycle.

 Passionate about leveraging AI to build cutting-edge solutions that drive the future of business innovation, from predictive analytics to intelligent automation. Deeply invested in staying at the forefront of AI-driven technologies, aiming to create transformative, next-gen applications that shape the future of industries. Motivated by the potential of AI to revolutionize operations and create intelligent systems that continuously adapt to evolving business needs.

**TECHNICAL SKILLS**

* **Programming/Coding:** Python (Pandas, NumPy, Scikit-learn, PySpark), R Programming, MATLAB, Java, Scala, C++, SAS/MACRO, SQL (Advanced SQL, T-SQL, PL/SQL), MySQL, PostgreSQL, Snowflake SQL, Databricks SQL, Shell Scripting, Alteryx, Git/GitHub, Unit Testing, Jupyter Notebooks, Google Colab
* **Advanced Excel:** Pivot Tables, VLOOKUP, HLOOKUP, INDEX-MATCH, Macros, Solver, What-if Analysis, Goal Seek, Data Tables, Scenario Manager, Regression Analysis, Simulation (using @RISK), Conditional Formatting, Dashboarding
* **Big Data Analytics & Visualization:** Apache Spark, Hadoop, Hive, HBase, MapReduce, Impala, Kafka, Airflow, Tableau, Power BI, QlikView, Looker, Google Data Studio, Kibana, Grafana, D3.js, OBIEE, Oracle Analytics Cloud (OAC/OAS)
* **Machine Learning & AI:** Machine Learning (Supervised/Unsupervised), Deep Learning, Generative AI, LLMs, Prompt Engineering, TensorFlow, PyTorch, Keras, OpenCV, XGBoost, LightGBM, CNNs, RNNs, LSTM, NLP (spaCy, NLTK, BERT, GPT), OCR, Computer Vision, Recommendation Systems, A/B Testing, Model Evaluation (ROC, Precision/Recall, F1 Score)
* **Data Engineering & Pipelines:** Databricks, Apache Airflow, Apache NiFi, ETL/ELT (SSIS, Talend, Informatica), Data Lakes, Data Warehousing (Redshift, Snowflake, BigQuery), Spark Structured Streaming, Flink, Kafka Streams, DBT
* **Cloud Platforms & DevOps:** AWS (S3, EC2, Lambda, Athena, Glue, EMR, SageMaker, QuickSight), Azure (Synapse, Data Factory, ML Studio), GCP (BigQuery, Vertex AI), Docker, Kubernetes, Jenkins, Terraform, EKS, ECS, GitHub Actions, DevSecOps, CI/CD
* **Natural Language Processing (NLP):** Tokenization, Lemmatization, Stemming, N-grams, Named Entity Recognition (NER), Text Classification, Sentiment Analysis, BERT, GPT, Word2Vec, TF-IDF, LDA Topic Modeling, Text Summarization
* **Time Series & Statistical Modeling:** ARIMA, SARIMA, VAR, VARMA, Exponential Smoothing, Holt-Winters, Prophet, PCA, Lasso, Ridge, Logistic Regression, ANOVA, MANOVA, Hypothesis Testing, MAPE, RMSE, MAE, TIC
* **Version Control & Collaboration:** Git, GitHub, GitLab, Bitbucket, Jira, Confluence, Agile/Scrum, VS Code, PyCharm
* **Certifications:** Google Analytics, Google Ads (AdWords), AWS Certified Data Analytics, AWS Machine Learning Specialty, SAS Base & Advanced, Microsoft Certified: Azure Data Scientist Associate, IBM Data Science Professional Certificate, NCFM – Financial Markets, IBM RFT, IBM RAD, Databricks Lakehouse

**PROFESSIONAL EXPERIENCE**

**Product Data Analyst, Capital One, McLean, Virginia Jan 2023-Present**

* **Designed and optimized scalable data pipelines using PySpark and Databricks SQL:**Partnered with data engineering teams to design and automate reporting systems tracking Month-over-Month metrics across publishing platforms (Mammoth and Spark Execution Engines). Leveraged Python, PySpark, and Databricks SQL to streamline data workflows, automate recurring reports, and eliminate 40 hours of manual work monthly enhancing operational efficiency and enabling strategic data-driven decisions.
* **Built centralized dashboards for KPI and OKR tracking across platforms:**
Developed comprehensive dashboards to consolidate data from multiple systems, enabling real-time monitoring of KPIs and OKRs for platform performance, partner onboarding, and product support. Automation of the data pipelines supporting these dashboards improved reporting accuracy and reduced monthly manual effort by 40 hours, while increasing transparency into customer lifecycle and retention trends.
* **Led product monitoring and KPI observability through cloud-based BI tools:**Architected and implemented a Data Observability dashboard for Lineage, EDQ, and DMA products using Databricks, Snowflake SQL, and AWS QuickSight. Ensured consistent metric tracking across disparate data sources, empowering leadership with real-time performance insights and supporting data-backed marketing and sales strategies.
* **Drove cross-functional analytics for business strategy alignment:**Collaborated with product, engineering, and operations teams to automate data pipelines and build interactive dashboards using Snowflake, Databricks, and QuickSight. Extracted and transformed data from APIs, Snowflake, and Amazon S3, delivering actionable insights that guided product development and business planning.
* **Standardized reporting practices and improved data reliability:**Developed intuitive BI dashboards and analytics frameworks to ensure consistency and integrity in KPI reporting across departments. Used Snowflake and QuickSight to provide stakeholders with trusted, self-service analytics tools that improved decision-making and promoted alignment across business units.
* **Enhanced segmentation and personalization through data-driven customer analysis:**Partnered with marketing and product teams to refine customer segmentation strategies by analyzing behavioral patterns and segment preferences. Used advanced analytics to inform targeting tactics, optimize conversion funnels, and personalize engagement efforts, improving campaign effectiveness and customer satisfaction.
* **Automated business operations with Databricks Job scheduling:**Led automation initiatives by scheduling data pipelines and reporting scripts using Databricks Jobs, replicating Airflow-like orchestration. Eliminated manual processes, increased operational reliability, and ensured seamless integration of automation into business workflows through effective documentation and change management.
* **Delivered enterprise-grade BI solutions for holistic performance monitoring:**Built and maintained dynamic BI dashboards that visualized end-to-end business operations. Introduced KPI Trees to help stakeholders identify root causes and track performance against strategic goals, enabling continuous process improvement and data-driven leadership.
* **Optimized advertising performance and ROI through partner data analytics:**Collaborated with external vendors to analyze advertising channel performance across digital platforms. Delivered insights that influenced media spend strategy and customer acquisition initiatives, maximizing ROI and aligning spend with high-performing segments and campaigns.

**Roles and Responsibilities**

* Design and implement scalable ETL pipelines using PySpark, SQL, and Databricks to automate data ingestion, transformation, and loading from diverse data sources (APIs, S3, Snowflake, SQL Server).
* Build and maintain data warehouses and data lakes on cloud platforms like AWS and Azure, ensuring structured storage, efficient querying, and compliance with data governance standards.
* Develop interactive and executive-ready BI dashboards using tools like AWS QuickSight, Tableau, and Power BI for real-time KPI tracking, OKR monitoring, and operational performance reporting.
* Architect and automate ML/AI pipelines for classification, clustering, and forecasting models using Python, Scikit-learn, TensorFlow, or PyTorch, and integrate them into production environments via REST APIs.
* Schedule and orchestrate workflows using Databricks Jobs or Airflow, managing dependencies, monitoring failures, and ensuring timely execution of recurring scripts and reports.
* Collaborate with cross-functional teams (product, marketing, finance, operations) to gather requirements, translate business goals into data models, and deliver analytical insights that drive strategy.
* Design data dictionaries and metadata documentation to standardize definitions, promote transparency, and foster self-service data access across business units.
* Work with DevOps and security teams to manage infrastructure as code, enforce data encryption, audit access controls, and ensure scalable, secure deployment of analytics solutions.
* Mentor junior data professionals and analysts, leading code reviews, knowledge-sharing sessions, and best practices in data modeling, visualization, and automation.
* Continuously evaluate and adopt emerging AI technologies, including LLMs and generative AI, to develop intelligent systems that automate decision-making and enhance customer experiences.

**Business Data Analytics, EXL Services Analytics, New York NY Dec 2021-Dec 2022**

* **Developed AI-powered media tagging and sponsor value estimation models:**
Designed and implemented a computer vision-based machine learning model using AWS S3, SageMaker, and Rekognition to detect sponsor brand logos, player identifier numbers, and unbranded white spaces in NFL game video and image footage. Achieved 94% accuracy in media tagging, enabling automated sponsor media value estimation and significantly enhancing sponsorship reporting capabilities.
* **Enabled privacy-compliant data collaboration through clean room architecture:**
Architected a secure data clean room platform to facilitate regulated data sharing between sponsors and partners. The solution supported anonymized, merged fan data analysis for AI/ML insights on NFL Owned & Operated platforms, enabling successful execution and performance measurement of sponsorship campaigns while adhering to data privacy standards.
* **Led end-to-end sponsorship analytics and machine learning initiatives:**
Owned and enhanced the entire NFL sponsorship lifecycle, including the development and optimization of machine learning models for content analytics and campaign performance. Delivered a unified Sponsor 360 strategy to drive measurable impact through predictive analytics and insight-driven decision-making.
* **Advanced audience segmentation with integrated data pipelines:**
Produced detailed audience profiling by integrating diverse fan data sources using Alteryx. Developed blind merge analytics techniques to enable high-level fan segmentation, empowering sponsors to effectively identify and target key audience segments for personalized engagement strategies.
* **Quantified sponsorship impact through data science-driven media attribution research:**
Conducted collaborative research leveraging fused audience datasets, including TV viewership, paid media, consumer behavior, and spend data. Applied advanced data science methodologies to attribute sponsor media investment to consumer outcomes, delivering actionable insights that optimized campaign ROI and sponsorship strategy.

**Roles and Responsibilities**

* Developed and deployed AI-driven computer vision models to automate media tagging and visual content analysis, significantly improving the accuracy and efficiency of value estimation and content classification.
* Designed and implemented secure data sharing environments using clean room architecture to enable privacy-compliant collaboration between business partners while ensuring compliance with data governance and regulatory standards.
* Led end-to-end development of analytics and machine learning solutions, including model training, optimization, and deployment, to drive strategic insights and support high-impact business decisions.
* Built scalable machine learning pipelines using cloud platforms and automation tools to support real-time data processing, predictive modeling, and business intelligence reporting.
* Performed advanced segmentation and profiling by integrating and analyzing large datasets to identify key user behaviors and optimize targeting and personalization strategies.
* Conducted attribution modeling and media impact analysis using statistical and machine learning techniques to measure the effectiveness of marketing efforts and guide investment decisions.
* Collaborated with cross-functional stakeholders including engineering, marketing, product, and compliance teams to deliver integrated analytics solutions that align with business objectives and data privacy requirements.
* Applied data science techniques such as logistic regression, clustering, time series forecasting, and uplift modeling to extract actionable insights and support campaign planning and optimization.
* Managed data integration and transformation workflows across cloud platforms and analytical tools, ensuring high data quality, consistency, and readiness for analysis and reporting.
* Presented analytical findings to business leaders and non-technical audiences, translating complex data insights into clear, strategic recommendations to support revenue growth and operational efficiency.

**Data Scientist, National Institutes of Health, Bethesda MD Feb 2019-Dec 2021**

* **Developed optimization models driving $2.2M in operational savings:**
Engineered a Chiller Decision Support System using mixed-integer non-linear programming (MINLP) models in MATLAB. Incorporated advanced optimization techniques—including Genetic Algorithms, Particle Swarm Optimization, and Differential Evolution—to reduce HVAC energy consumption, resulting in $2.2M in cost savings over 18 months for a large-scale industrial facility.
* **Built LSTM-based predictive models for energy load forecasting:**
Implemented and tuned Long Short-Term Memory (LSTM) neural networks using MATLAB’s Optimization Toolbox for campus-wide energy demand forecasting. Integrated 24/36-hour forecast outputs into real-time optimization systems, improving load prediction accuracy by 21% and enabling smarter energy scheduling.
* **Designed scalable data pipelines for industrial IoT analytics on AWS:**
Automated ingestion and processing of structured/unstructured time-series sensor data from OSIsoft PI System (PI Square, PI AF, PI Web API) using Python (PIAFSDK), PySpark, R, SQL, and MATLAB. Streamlined batch and real-time data processing pipelines into AWS (S3, Lambda, EC2), enabling continuous model training and real-time insights for anomaly detection and system optimization.
* **Performed advanced multivariate and predictive analytics at scale:**
Applied statistical and ML techniques—MANOVA, PCA, Lasso, Ridge regression, Random Forest, Gradient Boosting, and CNNs—to extract insights from complex multivariate datasets. Used Python (scikit-learn, TensorFlow, Keras) and R to build scalable ML models, reducing data analysis turnaround time by 35% and improving model explainability.
* **Time-series modeling and validation for operational forecasting:**
Built robust time-series models (ARIMA, SARIMA, VARMA) for sensor-based anomaly detection and predictive maintenance. Validated performance using RMSE, MAE, MAPE, and Theil’s Inequality Coefficient, achieving >90% forecasting accuracy across multiple use cases and supporting reliability improvements in operations.
* **Integrated ML and optimization workflows with AWS-based analytics platform:**
Deployed models and optimization pipelines in AWS (S3, EC2, Lambda), enabling end-to-end automation of data workflows and model outputs. Delivered real-time dashboards and alerts that supported operational decision-making and drove proactive maintenance strategies across energy systems.

**Roles and Responsibilities**

* Developed advanced optimization models using mixed-integer and non-linear programming (MINLP) techniques to drive operational efficiency and cost savings across industrial systems and processes.
* Implemented machine learning-based forecasting models, including LSTM neural networks, to predict resource demand and enable smarter scheduling and load balancing in large-scale operations.
* Engineered scalable data pipelines for real-time and batch processing of time-series sensor data from industrial IoT sources, ensuring reliable data flow for analytics and machine learning.
* Automated ingestion and transformation workflows by integrating diverse data systems using Python, PySpark, R, SQL, and MATLAB, enabling seamless end-to-end analytics in cloud environments.
* Applied multivariate statistical techniques (e.g., PCA, MANOVA, Lasso, Ridge) and predictive machine learning models (e.g., Random Forest, XGBoost, CNNs) to generate insights from high-dimensional datasets.
* Built robust time-series models (ARIMA, SARIMA, VARMA) to support anomaly detection, operational forecasting, and predictive maintenance strategies, improving system reliability and uptime.
* Deployed ML and optimization solutions to cloud platforms (e.g., AWS S3, Lambda, EC2), automating workflows and enabling real-time monitoring, alerting, and decision support dashboards.
* Performed model validation and performance tuning using industry-standard metrics (RMSE, MAE, MAPE, Theil's U), ensuring high accuracy, interpretability, and consistency across use cases.
* Collaborated with cross-disciplinary teams to define business problems, develop predictive and prescriptive models, and integrate AI-driven recommendations into operational decision-making.
* Streamlined industrial analytics processes through automation and integration, resulting in significant cost savings, enhanced energy efficiency, and improved visibility into system performance.

**Software Developer Analytics Intern, ATMECS Inc**, **Santa Clara, California**  **May 2018–Aug 2018**

* **Built predictive and classification models for business decision-making:**
Developed and fine-tuned a suite of machine learning models—including Linear and Logistic Regression, Decision Trees, SVM, XGBoost, Random Forests, KNN, and K-Means—for predictions, classifications, and clustering analysis. Utilized Python and R with libraries such as scikit-learn, TensorFlow, NumPy, SciPy, Matplotlib, and Seaborn to optimize model performance, delivering insights that supported data-driven decisions across product and marketing teams.
* **Delivered stakeholder-ready data visualizations and insights:**
Designed and presented complex time-series plots, scatter diagrams, histograms, bar charts, and geographic maps using Tableau Desktop to convey actionable findings across multiple levels of stakeholders. Enhanced cross-functional alignment by transforming quantitative outputs into intuitive dashboards that supported strategic initiatives.
* **Performed large-scale data extraction, transformation, and visualization:**
Executed web scraping projects using Python (BeautifulSoup, HTML parsing) to extract data from external sources and stored results in MongoDB. Applied data aggregations and transformations for analytical processing and built visualizations on AWS Kibana to track real-time data trends and KPIs.
* **Developed NLP pipelines for advanced text analysis and insight generation:**
Created an NLP application using Python and NLTK to analyze large-scale text data. Applied advanced preprocessing techniques including stemming, lemmatization, tokenization, n-grams, word embeddings, and BERT models to identify meaningful word relationships and extract sentiment and thematic trends for business analysis.
* **Built machine learning models for content success and customer segmentation:**
Engineered a Book Adaptations prediction model using a Random Forest classifier that achieved 75% accuracy in forecasting movie success. Developed a customer segmentation model based on transactional and behavioral data to identify high-value customer segments, enabling the creation of targeted retention strategies. Leveraged Python, R, and AWS Machine Learning Service for development, training, and deployment.

**Roles and Responsibilities**

* Developed and fine-tuned predictive and classification models using algorithms such as Linear/Logistic Regression, Decision Trees, SVM, Random Forests, XGBoost, KNN, and K-Means to drive data-informed decisions across business functions.
* Utilized programming languages and ML libraries including Python (scikit-learn, TensorFlow, NumPy, SciPy) and R to optimize machine learning models for performance, scalability, and interpretability.
* Created impactful visual analytics and dashboards using Tableau Desktop to communicate complex data insights through time-series graphs, scatter plots, histograms, and geospatial maps to business stakeholders and leadership teams.
* Executed data extraction and web scraping projects using Python tools (BeautifulSoup, HTML parsing) to collect external data, store it in MongoDB, and perform transformations for real-time visualization using AWS Kibana.
* Built and deployed NLP pipelines using Python and NLTK to analyze unstructured text data, applying techniques such as tokenization, stemming, lemmatization, n-grams, word embeddings, and BERT for sentiment analysis and trend identification.
* Engineered a machine learning model for content success prediction, achieving 75% accuracy in forecasting media performance based on historical and contextual data, supporting content investment decisions.
* Designed customer segmentation models based on behavioral and transactional data to identify high-value audiences and guide retention and personalization strategies across marketing and product initiatives.
* Leveraged cloud-based machine learning platforms such as AWS Machine Learning Services to train, evaluate, and deploy scalable models, ensuring production-ready solutions aligned with operational goals.
* Collaborated with cross-functional teams including marketing, analytics, and product management to embed machine learning insights into business workflows and improve outcome predictability.
* Maintained a rigorous focus on model validation and data quality, ensuring reliable, ethical, and unbiased outcomes in all analytical and AI/ML projects.

**Business Data Analyst, Kotak Mahindra Old Mutual Life Insurance Ltd, Mumbai, India Jun 2013 - Jul 2016**

* **Designed predictive modeling solutions to improve customer retention:**
Developed a supervised classification model using logistic regression in SAS 9.4 to predict policyholder lapse or surrender behavior. Achieved 69% model accuracy, which enabled targeted interventions and contributed to a 34% increase in customer retention. Delivered actionable insights to senior leadership and cross-functional teams through interactive dashboards, enabling data-driven decision-making.
* **Optimized data querying and manipulation in relational databases:**
Crafted complex SQL queries, stored procedures, triggers, and views in MySQL and SQL Server to support predictive modeling and reporting workflows. Created DML scripts for seamless data manipulation and designed relational databases with comprehensive data dictionaries. Leveraged advanced SQL techniques including window functions, joins, and aggregations to manage and analyze large datasets efficiently.

**Roles and Responsibilities**

* Developed and implemented predictive models using supervised machine learning techniques to identify customer behavior trends and support business retention strategies.
* Utilized statistical tools (e.g., SAS) to build and validate classification models, improving decision-making through data-driven insights.
* Communicated analytical findings to stakeholders through interactive dashboards and visualizations, facilitating strategic planning and performance improvement.
* Designed and optimized SQL queries, stored procedures, triggers, and views to support data processing and reporting workflows in relational databases (MySQL, SQL Server).
* Created and maintained database objects, data dictionaries, and DML scripts for efficient data handling and transformation.
* Applied advanced SQL techniques including window functions, joins, and aggregations to analyze large datasets and ensure data accuracy and performance.

 **EDUCATION**

**Master of Science in Business Analytics**  **2017- 2018**

Robert H. Smith School of Business, University of Maryland, College Park, MD

**Bachelor of Technology (BTech) in Computer Science and Engineering 2008- 2012**

Gandhi Institute of Technology and Management, Visakhapatnam, India