**Professional Summary:**

* Skilled **Data Scientist** and **Data Engineer** with 8+ years of experience in designing and implementing data-driven solutions across **GCP, and AWS** cloud platforms.
* Designed scalable ETL pipelines using **Google Cloud Storage (GCS), BigQuery, Dataproc, and Apache Airflow**, ensuring efficient data ingestion, transformation, and compliance with **HIPAA regulations**.
* Developed and deployed Machine Learning Algorithms using **BigQuery ML, TensorFlow, PyTorch, and XGBoost,** optimizing demand forecasting and inventory management, reducing forecasting errors by 30%.
* Containerized and deployed ML models with **Docker, Vertex AI, and Kubernetes**, integrating **FastAPI** for real-time inference, achieving 99.9% uptime and scalable performance.
* Implemented CI/CD pipelines using GitHub Actions, Cloud Build, and Vertex AIPipelines, automating model training, validation, and monitoring with **MLflow and Vertex AI** Model Monitoring.
* Leveraged Artificial Intelligence, Statistical Modeling,Finance Statistical Analysis, Statistical Inference and Advanced SQL to develop data-driven learning solutions and optimize cross-channel marketing strategies in AdTech, enhancing channel optimization and marketing science through knowledge sharing and project planning alongside software engineers and data scientists.
* Translated business requirements into analytical processes and code-driven solutions, delivering ad hoc insights and scalable models using Unity and advanced data science techniques.
* Expertise in data architecture, ETL processes, and data warehousing and delivering high-quality data solutions, Product Development,Unstructured DataEnsemble that drive business success.
* Expertise in creating and maintaining data pipelines for data processing and migration using Google Cloud Platform,and AWS cloud technologies.
* Extensive experience in designing, building, and managing data solutions on the **GCP and AWSplatform**.
* Hands-on experience working with AWS cloud in performing ETL operations, data processing, and pipeline creation using **AWS DMS, Glue, Kafka, S3, AWS lambda, Event Bridge, EMR, and EC2.**
* Defined policies using **AWS IAM** that specified the actions that users can perform on AWS resources.
* Expertise in handling and managing a large amount of data in cloud storage using **Redshift, and Snowflake.**
* Developed interactive visual dashboards, reports using **Tableau,Power BI, and Amazon QuickSight**.
* Expertise in implementing and managing robust and scalable **SQL** and **NoSQL** databases such as **MySQL, MongoDB, and SQL servers**.
* Experienced in all phases of Software Development Life Cycles, including **Agile and Scrum.**
* Effectively communicate complex data insights through written and verbal communication, adapting messaging for technical and non-technical stakeholders.
* Developed technical documentation, managed project timelines, designed and implemented machine learning models, and conducted simulations for predictive maintenance in machinery systems, focusing on bearings and gearing.
* Developed and deployed predictive models for fraud detection, anomaly detection, and healthcare compliance, leveraging LLMs, vector datastores, and Flask-based APIs for automation, while ensuring robust data governance, secure data exchange, and scalable data wrangling pipelines.
* Experienced in automating developing and deployment processes by implementing **CI/CD** using **Jenkins.**
* Experienced in orchestrating complex workflow, ensuring efficient and reliable data processing in implementing and synchronizing data pipelines using **Airflow.**

**TECHNICAL SKILLS:**

|  |  |
| --- | --- |
| **GCP Cloud** | Google Cloud Storage (GCS), BigQuery, Dataproc, PySpark, Pandas, Apache Airflow, Cloud Composer, TensorFlow, PyTorch, Scikit-learn, XGBoost, LSTM,K-Means, Docker, FastAPI, Kubernetes, Vertex AI, GitHub Actions, Cloud Build, Vertex AI, , MLflow, HIPAA Compliance. |
| **AWS Cloud** | S3, Redshift, Glue, RDS, Kinesis Streams, Kinesis Data Redshift Spectrum, EMR clusters, Lambda, SNS, Cloud Watch, Quick Sight, SageMaker, AWS Step Functions, Kinesis, , SageMaker Model Monitor, Grafana, CodeBuild, Terraform. |
| **Big Data Ecosystem** | HDFS, MapReduce, Hive, Pig, Sqoop, Cassandra, Apache Spark, Spark Streaming, HBase, YARN. |
| **Databases** | Mongo DB, My SQL, SQL Alchemy Oracle, DynamoDB, PostgreSQL, Maria DB, SQL, NoSQL Database-MongoDB, and Cosmos DB. |
| **Visualization and Reporting** | Tableau, PowerBI, Synapse BI, Quick Sights. |
| **Other Tools** | Kafka, Airflow, Snowflake, Git, Kubernetes, Jenkins, DataDog, |

**PROFESSIONAL EXPERIENCE**:

**Client: Cardinal Health, Remote. Jan 2023 - Present**

**Role: Data Scientist/ AI ML**

Developed and deployed AI-driven solutions for predictive analytics, supply chain optimization, logistics, and demand forecasting, replacing legacy systems with automated, scalable architectures that enhanced functionality. Improved operational efficiency, reduced forecasting errors, and accelerated speed to market while ensuring high data integrity, enabling real-time inference, and proactively monitoring model performance. Significantly improved accuracy and performance, providing a strategic competitive edge in the market and driving long-term business growth.

**System Design:**

* Developed and deployed **ML and Deep Learning** models for computer vision and predictive analytics using CNNs, LLMs, and advanced frameworks like TensorFlow, PyTorch, and Keras to improve supply chain efficiency, optimize inventory management, and enhance demand forecasting.
* Developed scalable code to fulfill dynamic business requirements, executed analytical processes for actionable insights, delivered adhoc analyses, and collaborated cross-functionally using Unity-driven data visualization frameworks.
* Extracted and processed large-scale healthcare datasets from **Google Cloud Storage (GCS)** using **BigQuery**, and **Dataproc**, leveraging **PySpark** and **Pandas** for scalable data transformation, data cleaning, data manipulationand data normalization to ensure high data quality and integrity with **HIPAA** regulations for analytical modeling.
* Developed and deployed machine learning models, including **XGBoost, LSTM, and Prophet** to accurately predict future demand, inventory levels, and capacity needs, reducing forecasting errors by 30**%.**
* Executed **train/test** splits using **Scikit-learn** & **TensorFlow**, applying **PCA**, feature scaling, and dimensionality reduction utilized **Pandas, NumPy,** and **PySpark** for optimized model performance.
* Designed, developed, and optimized ML models using **TensorFlow, PyTorch, and Scikit-learn,** implementing RandomForest, Gradient Boosting, and Deep Neural Networks for predictive analytics and optimization.
* Implemented K-Means, DBSCAN, and Causal Inference Models for supply chain optimization and customer/product segmentation, improving service efficiency by 20% and decision-making speed by 40%.
* Optimized model performance using**,**  hyperparameter tuning, feature selection, and cross-validation to enhance model accuracy and reliability.
* Containerized ML models with **Docker** and deployed on **Vertex AI and Kubernetes**, integrating **FastAPI** for real-time inference, achieving **99.9% model uptime** and scalable performance.
* **Designed and implemented end-to-end ML pipelines** by integrating data extraction, transformation, and model deployment with PySpark, Apache Airflow, Cloud Composer, and Vertex AI, optimizing ETL workflows, ensuring HIPAA-compliant data integrity, and achieving 99.9% model uptime with Kubernetes and Docker-based containerization.
* Designed and optimized **ETL pipelines** using **Apache Airflow and Cloud Composer** to automate data ingestion, transformation, and model retraining workflows.
* Designed interactive BI dashboards in **Tableau** leveraging EDA with Pandas Profiling, **Seaborn, and Matplotlib** to uncover patterns, detect anomalies, and drive data-driven decisions.
* Implemented **CI/CD pipeline** for automated model training, validation, and deployment using **GitHub Actions, Cloud Build, and Vertex AI Pipelines**, enhancing deployment efficiency.
* Continuously monitored ML models with **Vertex AI and MLflow**, ensuring optimal performance, timely drift detection, and automated retraining to maintain accuracy
* Developed AI-driven **business process automation** solutions, optimizing **analytical workflows, product delivery, and data collection** while enhancing **forecasting accuracy** and **operational efficiency** by 30%.
* Implemented AI-driven solutions for **supply chain optimization and logistics**, enhancing **predictive analytics, demand forecasting**, and **operational efficiency**, resulting in improved inventory management and faster delivery times.
* Designed and implemented A/B testing frameworks to evaluate AI-driven predictive analytics and forecasting models, enabling data-informed decision-making for supply chain optimization.

**Tool Stack**: Google Cloud Storage (GCS), BigQuery, Dataproc, PySpark, Pandas, Apache Airflow, Cloud Composer, TensorFlow, PyTorch, Scikit-learn, XGBoost, LSTM,K-Means, Docker, FastAPI, Kubernetes, Vertex AI, GitHub Actions, Databricks, Synapse, Cloud Build, Vertex AI, , MLflow, Grafana ,Tableau ,HIPAA Compliance.

**Client: Comcast**,**Phlidelphia, PA. Mar 2022 - Dec 2022**

**Job Role: Data Scientist/ AI ML**

Developed and deployed LLM-powered solutions to analyze customer transcripts, leveraging embeddings and clustering techniques to identify anomalies in customer behavior and patterns, operational inefficiencies, and irregularities impacting customer experience. Designed and implemented a data-driven framework to assess uncovering key insights into transfer efficiency and customer satisfaction, which directly shaped operational strategies. Delivered actionable insights that enhanced customer experience, reduced process inefficiencies, and supported data-driven decision-making at the organizational level.  
  
**System Design:**

* Developed and deployed **LLM models** using **AWS SageMaker**, applying embeddings, clustering techniques (K-Means, DBSCAN), and **NLP** techniques to detect operational inefficiencies and customer experience issues from call transcripts, with infrastructure provisioned using **Terraform and CloudFormation**.
* Extracted and processed large-scale customer transcripts stored in AWS **S3**, automating ingestion and preprocessing with **AWS Glue, Lambda, and Step Functions**, while streaming real-time data using **Kinesis** for immediate analysis.
* Built and optimized **LSTM**-based predictive models to forecast credit spend, churn likelihood, and customer satisfaction metrics, achieving a 25% reduction in churn prediction errors through advanced feature engineering, hyperparameter tuning, and cross-validation.
* Applied dimensionality reduction techniques (**PCA, t-SNE**), feature selection, and scaling methods using **Scikit-learn**, improving model interpretability and reducing processing overhead.
* Conducted comprehensive **exploratory data analysis** (EDA) using **Pandas** Profiling, **Seaborn, and Matplotlib**, uncovering patterns and anomalies to inform feature engineering and model selection strategies, with EDA insights stored in Elasticsearch/OpenSearch for historical analysis.
* Developed and deployed a data-driven framework in **SageMaker** Pipelines to assess and optimize call transfer policies, enhancing first-call resolution rates, with real-time metrics monitored via Grafana and CloudWatch dashboards.
* Containerized machine learning models using **Docker** and deployed to **SageMaker** endpoints, integrating with API Gateway and Lambda for low-latency real-time inference, while storing prediction logs and performance metrics in OpenSearch for trend analysis.
* Implemented continuous model monitoring using SageMaker Model Monitor and **CloudWatch**, ensuring early detection of data drift, concept drift, and performance degradation, triggering automated retraining workflows through Step Functions.
* Designed and orchestrated ETL pipelines using **AWS Glue** and **Apache Airflow**, automating data ingestion, transformation, and model retraining workflows, with infrastructure provisioned using Terraform and CloudFormation.
* Collaborated with cross-functional teams, including data engineers, product managers, and call center operations, ensuring model insights directly influenced operational strategies, with real-time insights visualized using **Grafana**.
* Developed and optimized geospatial data models and quality management systems, ensuring data accuracy, continuous improvement, and industry compliance while leveraging advanced data analysis techniques to enhance insights and decision-making.
* Implemented CI/CD pipelines using **CodePipeline** and **CodeBuild**, au1tomating the model training, validation, and deployment processes, ensuring seamless, reproducible deployments with infrastructure changes tracked using **Terraform** and **CloudFormation**.

**Tool Stack:** AWS SageMaker, AWS Glue, AWS Lambda, AWS Step Functions, AWS S3, AWS Kinesis, SageMaker Pipelines, SageMaker Model Monitor, CloudWatch, Elasticsearch/OpenSearch, Grafana, Docker, API Gateway, Apache Airflow, CodePipeline, CodeBuild, Terraform, CloudFormation, Scikit-learn, Pandas Profiling, Seaborn, Matplotlib, K-Means, DBSCAN, LSTM, PCA, t-SNE.

**Client: ETSY,Dallas,Texas. Jan 2021 - Feb 2022**

**Role: Data Scientist/ AI ML**

Developed and deployed recommendation and search optimization systems to replace legacy personalization models, enhancing functionality, flexibility, and self-service capabilities. Improved product discovery, user engagement, and conversion rates by implementing scalable deep learning and reinforcement learning-based ranking models. Optimized recommendation accuracy and diversity, accelerating speed to market while ensuring fairness and personalization in marketplace dynamics. Designed and automated end-to-end ML pipelines, enabling real-time inference, seamless data integration, and proactive model monitoring to drive business growth, maximize revenue, and provide a competitive advantage in e-commerce personalization.

**System Design:**

* Developed and deployed ML and Deep Learning models for personalized recommendations and search optimization using **CNNs, LLMs**, and advanced frameworks like **TensorFlow, PyTorch, and Keras** to enhance buyer-seller interactions, improve product discovery, and maximize revenue.
* Effectively utilized **Stream Analytics** to extract and store data within **ETL** workflows, resulting in the efficient and streamlined data processing.
* Extracted and processed large-scale **user-product interaction** datasets from **BigQuery and Dataproc**, leveraging **PySpark and Pandas** for data transformation, cleaning, and normalization to improve recommendation accuracy.
* Developed representation learning techniques using **Word2Vec, and BERT embeddings** to enhance product search, recommendations, and user intent predictions.
* Executed **train/test/validation splits** using **Scikit-learn & TensorFlow Data API**, applying **PCA, feature scaling, and dimensionality reduction** to enhance model performance.
* Developed **Graph Neural Networks (GNN)-based retrieval models** to enhance product discovery and user engagement.
* Built **deep learning** models for contextual search and personalized ranking, leveraging **CNNs, LLMs** (BERT, T5), and collaborative filtering techniques to enhance buyer-seller interactions.
* Developed machine learning pipelines by data mining, information retrieval, identifying and extracting key features and loading them into a feature bank for model training.
* Implemented **Reinforcement Learning** (RL) and **Multi-Armed Bandits** (MABs) for dynamic ranking, promotions, and ad placements, optimizing engagement and revenue.
* Optimized model performance using hyperparameter tuning, cross-validation, and feature selection with Optuna and Bayesian Optimization to improve recommendation relevance and diversity.
* **Containerized ML models** using **Docker** and deployed on **Vertex AI and Kubernetes**, integrating **FastAPI** for real-time inference, achieving **99.9% model uptime**.
* Designed and developed **scalable ETL pipelines** in **Oracle Autonomous Data Warehouse (ADW)** on **Oracle Cloud Infrastructure (OCI)**, optimizing data ingestion, transformation, and storage, ensuring high data accuracy, security, and compliance with industry best practices.
* Designed and implemented automated workflows using Airflow DAGs to seamlessly integrate data from multiple sources, including GCS and BigQuery,Kafka ensuring efficient ingestion, transformation, and storage in Parquet format within GCS buckets.
* Designed and implemented **CI/CD pipelines** for **automated model training, validation, and deployment** using **GitHub Actions, and Vertex AI Pipelines**.
* Monitored recommendation models using **Vertex AI** Model Monitoring and **MLflow**, tracking performance metrics, detecting model drift, and automating retraining workflows to ensure relevance, fairness, and optimal personalization.

**Tool Stack**: Google Cloud Storage (GCS), BigQuery, Dataproc, Vertex AI, Cloud Build, Cloud Composer, PySpark, Pandas, Apache Airflow, Scikit-learn, TensorFlow Data API, TensorFlow, PyTorch, XGBoost, DeepFM, Graph Neural Networks (GNNs), Reinforcement Learning (RL),Docker, FastAPI, Kubernetes, GitHub Actions, MLflow, Tableau, Google Data Studio, IntelliJ idea, Jupyter, Cloud Build, Cloud Sahale, Matplotlib,GitLab, Github,Excel.

**Client: IDBI federal life insurance co. ltd,Hyd,Inda**

**Role: Data Engineer**  **Aug 2018-Dec 2020**

**Description:**

Designed and developed scalable data pipelines to replace legacy ETL systems, enhancing flexibility and self-service capabilities. Optimized real-time data streaming, transformation, and storage for high data integrity and performance. Automated workflows and CI/CD pipelines to accelerate data processing and reduce operational overhead. Improved reporting accuracy and ensured timely data availability for analytics and business intelligence. Enabled data-driven decision-making, providing a competitive advantage in the insurance sector.

**System Design:**

* Developed and Designed **Kafka** clusters for real-time data streaming from **HDFS** to **AWS S3,** resulting in efficient and scalable data storage.
* Developed **Pyspark** scripts for data transformation and conversion and managed **EC2** instances for large-scale data processing in EMR.
* Designed and implemented data solutions using **Redshift** for loading the processed data to perform data analytics.
* Expert in using AWS Data Pipeline for automating data migration workflows, ensuring timely execution and effective use of resources.
* Ability to work with large and complex datasets and proficiency in using distributed file systems like **HDFS** and object stores like **S3.**
* Executed seamless migration of legacy Teradata workloads to Snowflake, using Airflow for orchestration, PySpark for data transformation, and AWS S3 for interim storage, resulting in improved pipeline efficiency and reduced infrastructure costs.
* Operations and applications are scheduled using **Airflow**.
* Conducted extensive data analysis using **Hadoop**, **Spark**, and **SQL**, providing key insights for business strategy and performance optimization.
* Handling large volumes of data in a scalable manner using **Hive** and **Scoop**, leveraging Hadoop's distributed file system to process large datasets in parallel.
* Effectively created and managed data pipelines using AWS to store and process data.
* Implemented **CI/CD** pipelines for Big Data applications using tools like **Jenkins and GitLab CI/CD.**
* Insights and reports are effectively generated utilizing visualization tools such as **Quick sight**.

**Tool Stack:** Kafka, Redshift, HDFS, Spark, SQL, Ec2, Airflow, AWS S3,Scala, Spark SQL, HDFS, Jenkins, GitLab.

**Client: IIH GLOBAL,HYD,IND.**

**Role: Python developer:**  **Feb 2017-July 2018**

**Description:**

Developed and deployed a scalable e-commerce application, enhancing functionality, flexibility, and self-service capabilities. Optimized performance, integrated SQL/NoSQL databases, and automated business processes to improve efficiency. Delivered high-quality web solutions, accelerating speed to market and driving business growth.

**System Design:**

* Developed Consumer-based applications and features using web technologies such as python and Django.
* Implemented Model View Control architecture in **Django Framework**.
* Expertise in building database designs using **SQL** and **NoSQL** databases.
* Created Business logic for planning and Tracking functions using **python.**
* Used python libraries like **NumPy, Beautiful Soup, Requests, and SQL Alchemy** for web application development.
* High-quality deliverables were achieved using **Agile**.
* Created applications using **RESTful** web services using MVC architecture.

**Tool Stack:** Python -NumPy, Beautiful Soup, Requests, SQL Alchemy, NoSQL, Django, HTML, CSS, Git, JavaScript, Linux.

**EDUCATION:**

* Master’s in Data Science (University of North Texas, Denton, TX)
* Bachelor’s in Computer Science - (GITAM University, Hyderabad, India )