Anagani Syam Prakash

LinkedIn Ph: +1 (561)-552-6049
sanagani2023@fau.edu | Address: 14401 S Military Tryl, Delray Beach, FL 33484

Detail-oriented Computer Science graduate with a strong foundation in data engineering, proficient in Python, SQL, and data visualization.Motivated and adaptable individual eager to apply strong technical and interpersonal skills in a dynamic role within the tech industry. Committed to continuous learning, innovation, and contributing effectively to organizational success. Passion for data pipelines, optimizing database performance, and implementing machine learning models in a Data Engineer role.

RELEVANT COURSEWORK

Software Engineering | Computer Science | Computer Networks | Web-based business development | Cloud computing | Computational Foundations of AI | Theory and implementation of database systems| Artificial Intelligence | Machine Learning | Sensor Networks & smart systems.

Masters Computer Science (Graduation) Jan 2023 - Dec 2024
Florida Atlantic University, Boca Raton, FL GPA 3.1/4.0
Bachelor of Engineering (UnderGraduation) **May 2022**
Sathyabama University, India GPA 7.9/10.0

TECHNICAL SKILLS

Data Engineer | Python Code | Core Java | DBMS | Microsoft SQL Server | Data Governance | Data Warehousing | ETL Pipelines | Powerbi | Data Access | Data Structures | Data Integrity | Data Visualization | Machine Learning Algorithm | Artificial Intelligence | Natural Language Processing | Statistical Analysis | Data Pipelines | Data Modeling | Data Storage | Data Reporting | Agile | Cloud Service | AWS | Azure | Snowflake | Scala | SSIS | GCP | Git |

SOFT SKILLS

Strategic Initiatives | Strategic Planning | Open Communication | Responsible | Prioritize | Write | Problem Solving | Commitment | Confident | Influence | Fidelity | Journals | Convert | Teamwork | Leading | Publish | Stem |

TOOLS AND TECHNOLOGIES

Programming Languages: C, Python(Numpy, Pandas, Matplotlib, Seaborn)

Databases: SQL, Sql Queries, Microsoft Office(Microsoft Word, Microsoft Tools, Microsoft Excel , Ms Powerpoint Ms Word,MTM)

Visualization Tools: Microsoft Power Bi, Tableau, MS-Excel(VBA, Pivot, Vlookup, Xlookup, Macros)

Cloud & DevOps: AWS(EC2, Lambda, DynamoDB), GCP (BigQuery), Azure, Docker, Kubernets

Big Data: Spark, Hadoop, Hive, Pig,Sqoop,Oozie

EXPERIENCE

Data Engineer – Tetra Tech . (Jul 2023 – May 2025)

• Designed and implemented data pipelines to integrate diverse data sources (e.g., satellite imagery, weather data, sensor networks, and social media) for real-time disaster monitoring and response.

• Developed and deployed interactive Power BI dashboards integrating real-time data from clinical trials, lab instrumentation, and patient monitoring systems, using Python and SQL to cleanse, transform, and model datasets, enabling data-driven decisions across pharmaceutical R&D and regulatory teams.

• Built predictive machine learning models using TensorFlow and PyTorch to forecast drug efficacy and patient outcomes, leveraging historical health records and trial data; results were visualized through dynamic Power BI reports to support evidence-based decision-making in drug development.

• Developed real-time data processing systems to analyze and visualize critical information during disasters, enabling faster decision-making for emergency response teams.

• Built predictive models using historical disaster data to forecast potential risks, such as flood zones, wildfire spread, or earthquake impacts, aiding in proactive disaster preparedness.

• Engineered geospatial data solutions to map disaster-affected areas, optimize resource allocation, and support evacuation planning.

• Ensured data accuracy and reliability by implementing data cleansing and validation processes for disaster-related datasets, improving the effectiveness of response strategies.

• Worked closely with government agencies, NGOs, and disaster response teams to provide actionable insights through data-driven dashboards and reports.

• Designed robust data backup and recovery systems to ensure continuity of operations and data availability during and after disasters.

Web Developer – Best Near Me - Remote. (Jun 2023 – Apr 2024)

• Designed and developed responsive websites using WordPress, customizing themes and templates to match brand identity and user experience requirements.

• Built custom pages using Elementor, WPBakery, and Gutenberg editors, ensuring cross-browser compatibility and mobile optimization.

• Integrated plugins for SEO, contact forms (e.g., WPForms), security (e.g., Wordfence), and performance optimization (e.g., caching tools like W3 Total Cache).

• Managed day-to-day website operations, including content updates, plugin/theme upgrades, performance monitoring, and backup management.

• Implemented SEO best practices using plugins like Yoast SEO and Rank Math to improve organic visibility and keyword ranking.

• Integrated Google Analytics and Google Search Console to monitor traffic, track user behavior, and identify optimization opportunities.

• Set up and maintained e-commerce functionality using WooCommerce, including product listings, payment gateways, and order management.

• Developed proof-of-concept data workflows on Google Cloud Platform (GCP) using BigQuery and Cloud Storage, evaluating performance and scalability for big data processing.

• Designed and implemented robust data pipelines to transfer and transform data from AWS DynamoDB into Hadoop.

Data Engineer – Helical IT Solutions Pvt Ltd . (Jan 2020 – Nov 2022)

• Microsoft Certified: PL-300 (Power BI Data Analyst Associate) with proven expertise in designing, building, and deploying interactive Power BI dashboards for business performance monitoring and data storytelling.

• Enabled impactful Power BI reporting by designing robust backend logic using SQL on Databricks, ensuring clean, aggregated datasets were readily available for dashboard visualization and business insights.

• Built and maintained efficient ETL pipelines using PySpark and SQL in Databricks, automating data ingestion and transformation from diverse external sources to support Power BI datasets and reporting layers.

• Developed SQL stored procedures and advanced queries to feed curated data models and reports in Power BI, optimizing for performance and reusability.

• Utilized Delta Lake features (Z-ordering, partitioning, caching) to enhance query speed and efficiency, resulting in faster refresh and load times for Power BI dashboards.

• Collaborated with data analysts and BI developers to define data models and schemas tailored for Power BI consumption, enabling self-service reporting and ad hoc analysis.

• Facilitated secure access to data for Power BI users through role-based access control (RBAC) and managed permissions within Databricks, ensuring data governance and compliance.

• Used Databricks notebooks for iterative development, testing, and validation of datasets ultimately used in Power BI visualizations.

• Supported integration between backend data services and front-end tools, including REST APIs, to enable dynamic, real-time Power BI dashboards.

Data & ML Engineer – U Future Tech. (Feb 2018- Jan 2020)

• Collaborated with cross-functional teams to design and scale robust data pipelines and architectures for large-scale datasets including text and 3D geometric data.

• Built and optimized deep learning models (CNNs, Transformers) using frameworks such as PyTorch and TensorFlow, ensuring production-ready deployment and reproducibility.

• Developed and maintained scalable ETL and ML pipelines using Spark, PyTorch Lightning, and Ray, enabling efficient data preprocessing, transformation, and model training.

• Hands-on experience working with Large Language Models (LLMs), embedding techniques, vector databases, and RAG systems for scalable content understanding.

• Applied computational geometry techniques and 3D data modeling to support advanced analytics in spatial and geometric datasets.

• Integrated cloud-native solutions across AWS (e.g., S3, Lambda, DynamoDB, SageMaker) and Azure to ensure scalable and secure model deployment environments.

• Conducted model experimentation and hyperparameter tuning, evaluated performance metrics, and implemented optimization techniques for improved latency and accuracy.

• Ensured model and data pipeline reliability through robust validation, monitoring, and recovery processes in real-time, mission-critical scenarios.

• Proficient in translating theoretical ML concepts into scalable, practical implementations with clean, well-documented code and architecture.

• Experienced in Agile methodologies, participating in sprint planning, requirement analysis, and continuous delivery of ML features and insights.

• Delivered data-driven insights through visualization tools like Power BI and Tableau, and communicated technical findings to non-technical stakeholders.

• Version-controlled data and models using Git and maintained reproducibility across ML workflows and experiment tracking platforms.

• Experienced in distributed computing using Apache Spark and Hadoop to scale ML training and analytics for big data environments.

• Proven ability to develop production-grade machine learning systems in fast-paced, cross-disciplinary environments involving AEC and healthcare domains.

• Demonstrated knowledge in data governance, data security, and cloud-based infrastructure management to ensure compliance and data integrity.

• Skilled in advanced SQL, query optimization, and tuning techniques for large RDBMS deployments, ensuring performance efficiency in cloud environments.

• Proficient in cloud deployments and CI/CD, with expertise in AWS services like RDS, Aurora, and Redshift, as well as Infrastructure as Code (IaC) using Terraform.

PROJECTS

Healthcare Cost Optimization and Patient Outcome Analysis Using Real-World Claims Data:

• Designed and implemented scalable ETL pipelines using Azure Data Factory and Python to process high-volume healthcare claims data, supporting real-time analytics and ML model integration.

• Applied deep learning and statistical models to forecast patient outcomes and treatment efficacy, enabling cost optimization and risk stratification across healthcare networks.

• Modeled healthcare data using star-schema and Lakehouse architecture, optimizing for fast querying and ML readiness in Azure Synapse and Power BI environments.

• Collaborated with data scientists to integrate predictive models into cloud workflows, ensuring scalable deployment and reproducibility through MLOps best practices.

• Automated data validation, lineage tracking, and anomaly detection processes to maintain high data quality and compliance (HIPAA), ensuring trustworthy insights.

• Visualized key healthcare performance metrics using interactive Power BI dashboards, aiding clinical and operational decision-making across R&D and regulatory teams.

• Tuned model hyperparameters and performed cross-validation to enhance predictive accuracy, leveraging tools like PyTorch and TensorFlow for model experimentation.

• Implemented access controls and data governance policies within Azure, aligning with security and compliance frameworks for sensitive healthcare data.

• Demonstrated agile collaboration and communication, translating complex ML insights into business-driven outcomes for stakeholders in clinical domains.

E-Commerce Product Recommendation System:

• Designed and implemented a collaborative filtering-based recommendation system for an e-commerce platform.

• Ingested large-scale product transaction data from MySQL into HDFS using Sqoop.

• Processed and cleaned data using Spark and Hive, ensuring structured storage.

• Developed a Recommendation Model using Spark MLlib (Collaborative Filtering - ALS).

• Built a REST API using Flask to serve personalized product recommendations to users.

• Integrated the system with Tableau to visualize customer preferences and product trends.

Financial Analytics System for Profit and Loss Management :

• I developed a database system to monitor and analyze a company's financial performance. We created structured tables for accounts (revenue and expenses).

• Your Financial Analytics System effectively uses structured databases and advanced SQL queries to track profit, expenses, and revenue trends, making it a robust financial analysis tool.

• Enhancing it with Power BI dashboards for real-time visualization, including dynamic KPIs, drill-down reports, and trend forecasting, can improve decision-making.

• Consider adding automation via scheduled stored procedures, predictive analytics (ARIMA, ML models for forecasting), and integration with cloud databases (Azure, AWS RDS) for scalability and real-time insights.

Scalable Video Streaming Platform Using AWS and CloudFront

• In order to provide safe storage, adaptive streaming, and worldwide scalability, your scalable video streaming platform efficiently makes use of AWS services (S3, MediaConvert, CloudFront, and Elastic Load Balancer).

• Performance and cost effectiveness can be further maximized by enhancing it with AWS Lambda for serverless processing, DynamoDB for metadata storage, and Auto Scaling for dynamic resource allocation.

• To make the platform more intelligent, responsive, and user-focused, think about including AI-based video recommendations, real-time analytics with Amazon Kinesis, and edge computing for ultra-low latency.

Sales Performance Dashboard in Power BI

• Top Product Contribution: Identified that the top 10 products contributed to 60% of total revenue, with an average sales growth of 15% month-over-month. Focusing marketing efforts on these test products could increase overall revenue by 10%.

• Regional Sales Disparity: Analysis showed that the West region underperformed by 20% compared to other regions, with a declining sales trend. Redirecting resources and sales strategies to this region could boost sales by 12%.

• Customer Segment Profitability: Discovered that premium customers, though only 30% of the customer base, accounted for 55% of total profits. Targeting similar high-value segments could increase profit margins by 8%..

Social Media Crime Detection using Machine Learning :

• My project focused on Crime Detection in Social Media using natural language processing and machine learning. The goal was to identify and classify potential crime-related content from social media posts, such as hate speech, harassment, or threats.

• Using TF-IDF and models like Random Forest, Gradient Boosting, and Logistic Regression, your project successfully uses NLP and ML for crime detection

• Accuracy and contextual comprehension can be increased by enhancing it with deep learning (BERT, LSTMs), real-time data (Twitter API), and feature engineering (Word2Vec, NER).

• It can be made more effective and scalable by including imbalanced data processing (SMOTE), multi-class categorization (hate speech, threats, cyberbullying), and deployment via Flask/FastAPI.