

# VENKATA BABU ALAPATI

Detroit, MI | 313-398-7709 | valapati01@gmail.com | LinkedIn: venkat-alapati | github: venkat-a | https://alapati.io

## Data Scientist & MLOps Specialist | Skilled in Data Engineering, Machine Learning, Cloud Infrastructure, and Predictive Analytics

Highly skilled Data Scientist with extensive experience in designing, implementing, and scaling machine learning models, MLOps pipelines, and cloud infrastructure on AWS and Azure platforms. Proven ability to drive operational efficiency, improve financial models, and enhance predictive maintenance systems through advanced data processing and automation frameworks. Expertise in statistical analysis, time-series forecasting, and large-scale data engineering, with a strong focus on optimizing costs and ensuring security compliance. Adept at leveraging cutting-edge technologies such as Generative AI, NLP, and container management to deliver impactful business insights and solutions. Committed to continuous improvement and innovation in fast-paced, complex environments.

### AREAS OF EXPERTISE

**Data Science:** Hypothesis Development, Machine Learning Algorithms, Deep Learning, Generative AI, Natural Language Processing (NLP), Predictive Analytics, Data Visualization, Business Intelligence, Statistical Modeling, Time Series Forecasting

**Data Engineering:** Data Pipelines, SQL, Python, Feature Engineering, Query Optimization, Real-Time Data Streaming, Data Aggregation

**Statistical & Quantitative Analysis:** Exploratory Data Analysis (EDA), Hypothesis Testing, Predictive Modeling, Causal Inference, Constrained Optimization, Statistical Distributions, Regression Analysis, Cost/Profit equations, Stochastic Heuristics

**Cloud Technologies:** AWS, Azure, Google Cloud Platform (GCP), Snowflake, Heroku, CLIs, Configuration Files (AWS, Azure), YAML

**Technical Documentation:** Requirements Analysis, Data Governance, Analytic reports, Base Models, and Code Templates

### PROFESSIONAL EXPERIENCE

Data Scientist | ADP – Full Time, Detroit, USA

Sep 2024 - Present

- Engineered multi-page KPI dashboards with TensorFlow Edge-AI models, utilizing SQL and Python for quantization, pruning, and visualization, reducing model size by 40% for scalable deployment
- Optimized Edge-AI TensorFlow models using assembly language, eliminating loops and implementing buffering pipelines to reduce inference time by 30% for real-time applications
- Implemented deployment workflows for pre-trained models from the Arm ML Zoo onto Ethos-U55 NPUs using AWS SageMaker, ensuring seamless infrastructure scalability
- Streamlined CI/CD pipelines using GitHub Actions, AWS Code Pipeline, and CloudWatch, reducing deployment timelines by 24% and improving security with IAM and VPC configurations

Data Scientist | Coders Data LLC – Full Time, Winona, USA

Jun 2024 – Aug 2024

- Designed and deployed Python-based CECL and CCAR models, integrating SAS and Alteryx to improve accuracy by 30%, streamlining financial risk assessment workflows
- Conducted causal analysis and Monte Carlo simulations, enabling data-driven financial decision-making and enhancing predictive capabilities
- Reduced unstructured data processing times by 40% through custom NER models with PyTorch and fine-tuned BERT models
- Performed statistical conformity analysis for \$5.5M in transactions using Benford's Law, maintaining deviations under 5% and optimizing credit limit validation processes
- Created Alteryx dashboards to reduce discrepancies by 30% and enhance \$8.1M credit limits and sales data validation

Data Scientist | Dash Metrix's Inc – Full Time, Delaware, USA

Oct 2023 - Jun 2024

- Developed time-series forecasting models (LSTM, ARIMA) on causal-inference data for macroeconomic indicators, improving stock price prediction accuracy by 20%
- Automated financial anomaly detection workflows with One-Class SVM, auditing 500+ transactions and reducing discrepancies by 30%.
- Optimized SQL queries with advanced aggregation techniques (ROLLUP, CUBE), improving NLP predictive model pipelines and reducing runtime by 40%.
- Improved patient risk stratification by fine-tuning Random Forest and XGBoost models, normalizing data, and optimizing SQL queries, enhancing accuracy by 30%.

Resident AI | Apziva – Remote, Part Time, Turkey

Oct 2023 - Jun 2024

- Developed document digitization models (ResNet50, EfficientNetB0) in SageMaker, fine-tuned hyperparameters to achieve 88% accuracy, outperforming base models by 17%.
- Applied SHAP and LIME on Random Forest features, reducing features by 70%, and optimized PCA, KMeans, and SMOTE to boost efficiency by 30%.
- Enhanced model performance with OneHotEncoder and precision-recall analysis, boosting F1 scores by 25%.
- Built NLP-driven predictive candidate ranking systems with LightGBM and BERT embeddings, achieving top 1% rank precision
- Enhanced ARIMA models with time series decomposition, seasonal components, and feature scaling, using AWS Data Wrangler to boost model predictive power by 30%.
- Visualized predictions and metrics with Matplotlib and Seaborn, improving interpretability by 25% for actionable insights.

**Summer Practicum Intern** | One Magnify – Full Time, Detroit, MI

**May 2023 - Aug 2023**

- Developed multi-platform KPI dashboards with Streamlit, SQL, and Python, reducing load times by 67% and enabling real-time updates via CLIs.
- Optimized CLI configurations and resolved data type issues in Snowflake, consolidating data from 8 sources and reducing dashboard access to 10 seconds.
- Engineered a time travel tool with causal time-series analysis, resolving timestamps for 245-day predictions, exceeding Snowflake limitations.
- Migrated Power BI, Tableau, and Alteryx dashboards to Streamlit, utilizing SQL and Python, cutting costs by 40% and improving efficiency.

**Software Engineer** | Danlaw Inc – Full Time, Novi, MI

**Jan 2023 - May 2023**

- Developed predictive ECU calibration models with Python and Git version control, automating CANalyzer configuration and test log workflows, enhancing accuracy by 15%.
- Led implementation of vehicle-specific ECU calibration projects, reducing calibration time by 20% and improving accuracy by 15%.

## FOUNDATIONAL EXPERIENCE

**Senior Software Engineer** | AMD India Development Center – Bangalore, India

**Software Developer** | Satyam Computers Limited - Hyderabad, India

**Technical Engineer** | General Electric India Innovation Center - Hyderabad, India

**R & D Engineer** | HBL Power Systems Limited - Hyderabad, India

## EDUCATION

**Master of Science** | **Data Science & Business Analytics**

Wayne State University- Mike Ilitch School of Business – Detroit, MI

**Master of Technology** | **Embedded Systems**

JNTU College of Engineering – Hyderabad, India

## CERTIFICATIONS

AWS - Practical Data Science | Oracle Certified Architect | MITx - Computational Programming with Python | LinkedIn - Scientific Data Science | MIT-CTL - Micro Masters in Supply Chain Management

## TECHNICAL PROFICIENCIES

**Programming Languages:** Python, R, SQL, C++, Shell Scripting, SAS, JavaScript, HTML, CSS

**Data Science Libraries:** Scikit-learn, TensorFlow, PyTorch, XGBoost, NumPy, Pandas, SHAP, Lime.

**Cloud & Infrastructure:** AWS (VPC, IAM, EC2, S3, Lambda, Sage Maker, Glue, Astana), Azure (AKS, ML Studio, AKS, ACI, DevOps)

**Tools & Platforms:** Jenkins, Docker, Kubernetes, Terraform, Git, IDEA, Alteryx, Excel Solver, MS Office, Confluence

**Visualization & Reporting:** Python (Matplotlib, Seaborn, Dash), Streamlit, Tableau, Power BI, SAP Analytics, Full stack Development

**CI/CD & MLOps:** GitLab CI/CD, AWS CodePipeline, AWS CodeBuild, AWS CloudWatch, Azure DevOps, Jenkins, Terraform

## PROJECTS

**Real-Time ETL & Predictive KPI Dashboards:** Architected real-time ETL pipelines for Snowflake (15M+ records, 25% faster) and multi-page predictive KPI dashboards with Streamlit, reducing visualization load times to 10s and automating workflows with CLIs

**Demand & Revenue Analysis (GB Bikes):** Boosted sales by 10%, cut attrition by 15%, and reduced inventory costs by 20% through advanced demand forecasting and time series analysis.