CHINMAYEE GORUGONTHU

Clarksville, Tennessee | +1(737) -288-6996 | chinnu.gvp@gmail.com | Linkedin | Github

SUMMARY

Versatile data professional skilled in the complete project lifecycle, from automated data pipeline construction to advanced model deployment. Leverages Python, SQL, TensorFlow, and Power BI to build classification models with 91% precision, enhance forecast accuracy, and support critical business functions. Consistently translates complex data challenges into measurable improvements in efficiency and strategic growth.

EDUCATION

University of Texas at Arlington - Texas, USA **August 2021 - May2023**
Masters in Data Science **-** 3.70 / 4 GPA

G. Narayanamma Institute of Technology - Hyderabad, India **August 2017 – May 2021**
Bachelors in Electronics and Communication Engineering – 7.8/10 GPA

WORK EXPERIENCE

CDE Lightband | Tennessee, USA | Data Scientist **Aug 2023 - Present**

- Provided foundational data analysis for a utility-wide Cost of Service Study by extracting, cleansing, and validating billing determinants and hourly load profiles for customers from AMI and SQL database systems. Developed automated Python scripts to aggregate time-series data, ensuring the accuracy and defensibility of the cost allocation model. Automated the generation of monthly Excel-based reports using Python, reducing manual preparation time by 40%.
- Reduced the forecast error (MAPE) of a deep learning load forecasting model by 10% through targeted feature engineering and hyperparameter tuning. Identified and incorporated new predictive features and optimized the deep neural network's architecture (TensorFlow, Keras) and dynamic scaling factor.
- Engineered and evaluated deep learning and ensemble models for a complex time-series classification task on AMI data. Addressed data scarcity and imbalance by generating synthetic data, leading to a 30% improvement in model recall. Validated the final model using a confusion matrix, achieving a precision of 91%, which provided a scalable method for identifying potential energy loss.
- Identified a 12% potential uplift in broadband market penetration by building a predictive regression model to correlate KPIs with market performance. Developed and delivered a Power BI report that visualized the gap between actual and predicted market share, providing a data-driven case for new strategic initiatives.
- Built a data pipeline using Python to extract data from the API Server. Utilized Pandas for in-memory transformations and loaded the processed data into a PostgreSQL database. Orchestrated the workflow to run every day and created a dataset of over million datapoints for back testing trading strategies.
- Delivered a projected 10% reduction in peak energy demand by developing a customer segmentation model in Python. The model used Scikit-learn (K-Means) on consumption data extracted from SQL Server using pandas and SQLAlchemy, successfully identifying distinct usage profiles to guide new energy efficiency programs.

Spyry Technologies |Bengaluru, India | Intern **May 2019 - September 2019**

- Developed and tested an unsupervised anomaly detection model using Isolation Forest and Scikit-learn to identify previously undetected C2 (Command and Control) communication patterns in massive network traffic logs. The model contributed to a 20% reduction in mean-time-to-detect (MTTD) for a critical threat vector, directly enhancing the product's core capabilities.
- Engineered a data processing pipeline using Python and pandas to parse, clean, and structure raw endpoint security logs from multiple sources. This automated the feature extraction process, making the data readily available for threat modeling and reducing data preparation time for the security research team by 75%.
- Analyzed a dataset of 30,000 phishing attempts using Natural Language Processing (NLP) techniques to extract key features predictive of malicious intent. Collaborated with the security engineering team to present findings, which were directly incorporated into the product's detection rules, improving the capture rate of new phishing campaigns by 15%.

SKILLS

- **Machine Learning**: Supervised Learning (**Regression**, Classification), Unsupervised Learning (**Clustering**, Dimensionality Reduction), Reinforcement Learning, **NLP** (Natural Language Processing), Deep Learning (**TensorFlow, Keras, PyTorch**), Model Evaluation & Validation, Feature Engineering, Time Series Analysis,
- **Programming**: **Python** (Pandas, NumPy, Scikit-learn, SciPy), **R (**dplyr, ggplot2), **SQL** (MySQL, PostgreSQL, SQL Server),
- **Statistical Analysis**: Hypothesis Testing, A/B Testing, Experimental Design, Statistical Modeling, Probability, Bayesian Statistics, Monte Carlo Simulation
- **Data Tools & Platforms**: Apache Spark, Hadoop (HDFS, MapReduce), Apache Kafka, **Tableau**, **PowerBI**, Qlik Sense, Matplotlib, Seaborn, **Excel** (advanced), Jupyter Notebooks, Airflow, Athena
- **Cloud Platforms:** Amazon Web Services (**AWS**: S3, EC2, SageMaker, EMR, Redshift),
- **Certifications:** AWS Certified Machine Learning Specialty