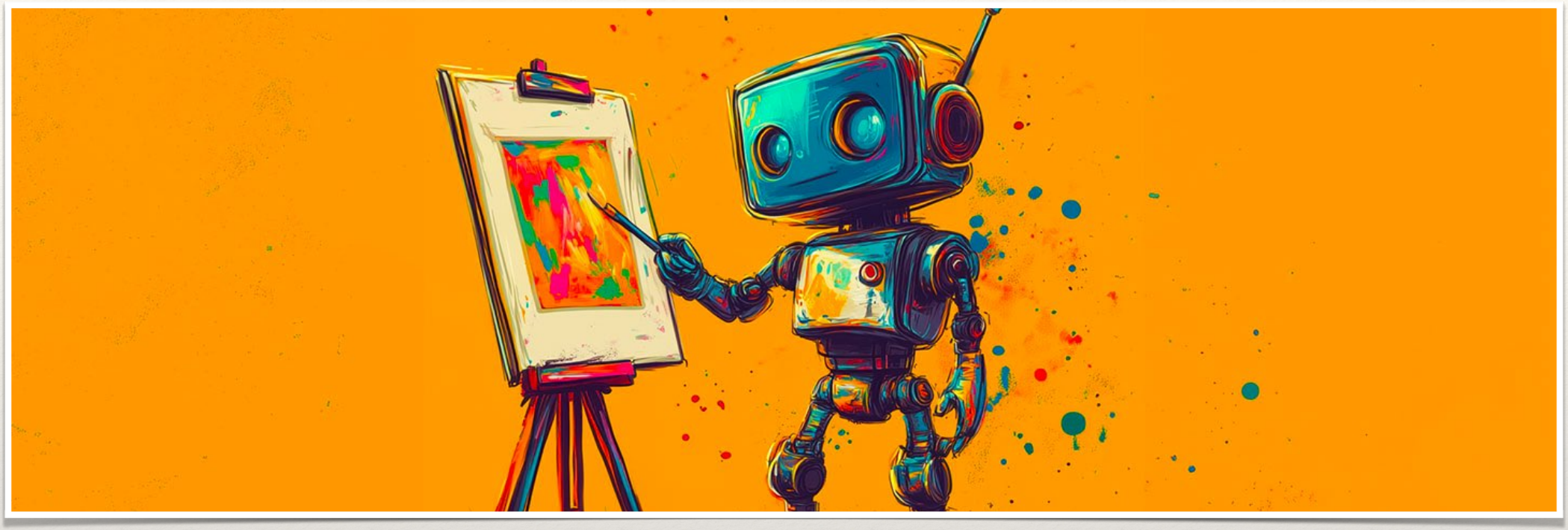

Gen AI For Observability

- R. Prabhas Reddy
- B. Harshath
- B. Abhishek

Introduction

In this project, we propose an **ensemble approach** that combines the strengths of three different models—**ARIMA**, **Prophet**, and **Isolation Forest**—to perform anomaly detection on time series data. Each model has unique capabilities: ARIMA captures linear trends and short-term dependencies in time series, Prophet handles seasonality and events like holidays, and Isolation Forest detects anomalies by isolating outliers in high-dimensional data. By leveraging an ensemble of these models, we aim to build a more robust and accurate system for detecting anomalies, ensuring that no single model's limitations affect the overall performance. The ensemble approach allows us to benefit from the complementary strengths of these models. ARIMA and Prophet provide reliable forecasts for time series with predictable patterns, while Isolation Forest excels in detecting rare and unexpected events that might go unnoticed by traditional forecasting methods. By integrating their predictions, we can create a unified anomaly detection system that not only adapts to various data characteristics such as seasonality, trends, and outliers but also ensures greater accuracy and reliability in identifying anomalies. This ensemble method enhances the robustness of monitoring systems, allowing for early detection of sensor malfunctions, environmental hazards, or other irregularities.

ARIMA(Auto-Regressive Integrated Moving Average)

The ARIMA model is a popular statistical approach used for forecasting time series data. It combines three key components: Auto-Regressive (AR), Integrated (I), and Moving Average (MA). The **AR** part models the relationship between current and past values, while the **MA** component captures the influence of past forecast errors. The **Integrated (I)** part involves differencing the time series to make it stationary, which is a requirement for ARIMA to work effectively. The model is represented by three parameters—p (AR order), d (differencing order), and q (MA order).

ADVANTAGES :

- Accurate for Short-Term Forecasting
- Effective for Stationary Data
- Widely Used and Well-Understood

PROPHET

The Prophet model, developed by Facebook, is designed to handle time series data that contain missing values, outliers, and seasonality. It models the data as a combination of a trend function, seasonal components, and holiday effects. Prophet is particularly suitable for forecasting business metrics like website traffic or sales, especially when the data exhibits daily, weekly, or yearly seasonal patterns. A significant advantage of Prophet is that it requires minimal tuning, making it accessible even to non-experts. Users can easily incorporate holiday effects or custom seasonal events into the model.

ADVANTAGES :

- Handles Missing Data and Outliers
- Built-in Seasonality and Holiday Handling
- Good for Business Applications

ISOLATION FOREST

The Isolation Forest is a machine learning algorithm that excels at detecting anomalies or outliers in data. It works by isolating data points through random partitioning. Since anomalies differ from normal data points, they tend to get isolated quickly, resulting in shorter paths within the tree structure. This makes the Isolation Forest particularly effective for high-dimensional datasets. It is commonly used in applications like fraud detection, network intrusion detection, and defect detection in manufacturing. One of its key strengths is efficiency, as it requires fewer computational resources than other anomaly detection methods.

ADVANTAGES :

- Efficient for Large and High-Dimensional Data
- Fast and Computationally Efficient
- Works Well with Streaming Data

Conclusion

In conclusion, this project successfully developed an **ensemble anomaly detection model** by combining ARIMA, Prophet, and Isolation Forest to monitor time series data of IOT devices. The ensemble leveraged ARIMA's ability to capture linear trends, Prophet's strength in handling seasonality, and Isolation Forest's capability to detect outliers, resulting in a more robust and accurate detection system. This approach effectively addressed the limitations of individual models, providing comprehensive anomaly detection across a wide range of scenarios. The project demonstrated the advantages of using an ensemble technique for time series anomaly detection, making it a powerful solution for real-time monitoring in various applications.

Thank You!