Capstone 3: Web Traffic Forecasting Project

Humayra Azra

Reason for Project

- Web traffic prediction is crucial for managing resources and improving user engagement.
- The challenge lies in capturing trends, seasonality, and anomalies in traffic data.
- Predicting traffic patterns accurately is essential for decision-making in marketing, content, and infrastructure.

Project Overview

- Objective: Forecast future web traffic based on historical data.

- Techniques: Time-series analysis, EDA, and machine learning.

- Methods: Random Forest and XGBoost models.

- Goal: To provide accurate, actionable insights into web traffic patterns.

Target Audience

- Business decision-makers who rely on web traffic data to inform marketing strategies.

- Data scientists and analysts working with web traffic data.

- Web development and infrastructure teams optimizing resource allocation based on traffic.

What We Can Monitor

- Web traffic trends (upward or downward).
- Seasonal patterns and peak traffic periods.
- Anomalies or sudden spikes in traffic.
- User behavior and engagement over time.
- Resource usage and optimization based on traffic predictions.

Significant Findings

- Web traffic follows predictable seasonal cycles.
- Strong correlation between past traffic patterns and future behavior.
- Models such as Random Forest and XGBoost achieved 94% accuracy in classification tasks.
- Seasonal decomposition revealed the presence of non-linear trends in the data.

Machine Learning Model

- Random Forest and XGBoost used to classify traffic as high or low.

- GridSearchCV was used for hyperparameter tuning.

- Both models demonstrated high accuracy (94%) and robust performance in predicting web traffic categories.

Future Improvements

- Incorporate lagged features for capturing more short-term dependencies.

- Integrate event-based data, such as holidays and marketing campaigns, to improve accuracy.

- Experiment with deep learning models like LSTM for capturing non-linear dependencies.

- Improve model robustness with additional data preprocessing techniques.

Next Steps

- Finalize model selection based on business needs.
- Deploy the selected model for real-time web traffic forecasting.
- Monitor model performance and fine-tune as new data becomes available.
- Provide actionable insights to stakeholders for improved decision-making.