Operational Longitudinal Analytics Validation Engine (OLAVE) for the National Football League (NFL): Internship at CAHSAA AT&T

Vasudevan Govardhanen
Science & Engineering
Manalapan High School
Englishtown, NJ
426vgovardhanen@frhsd.com

Siddharth Kedharnath
Science & Engineering
Manalapan High School
Englishtown, NJ
426skedharnath@frhsd.com

Abstract

OLAVE-NFL (Operational Longitudinal Analytics Validation Engine for the NFL) is a sports analytics platform designed to evaluate and visualize contextual NFL play-by-play data using predictive modeling and advanced statistical metrics. Current sports analytics systems often focus primarily on traditional player statistics, limiting deeper situational analysis. Our project addresses this by integrating play-by-play data from the nflfastR repository with machine learning models and interactive visualizations to quantify offensive efficiency and game-state value. Using over 100,000 NFL plays from the 2021–2023 seasons, we engineered features including down, distance, field position, score differential, and time remaining to calculate metrics such as Expected Points Added (EPA), Win Probability (WP), Success Rate, Red Zone Efficiency, and Drive-Level Efficiency. We implemented gradient boosting models through XGBoost to estimate situational success probabilities and contextual play value. The resulting analytics pipeline combines R, Python, pandas, Plotly, FastAPI, and PostgreSQL to support both predictive modeling and interactive visualization. OLAVE-NFL aims to provide coaches, analysts, broadcasters, and fans with a more comprehensive understanding of decision-making and efficiency in professional football.

Index Terms

AT&T, internship, CAHSAA, National Football League, NFL, sports analytics, play-by-play, statistics, nflfastR, R, Python, pandas, Plotly, FastAPI, PostgreSQL, predictive modeling, interactive visualization