NFL Gridiron Guru - The Ultimate Team NFL Matchup Engine

NFL Game Search Tool for Detailed Insights

By: Matthew Gorena and Sean Lee

Project Overview

Objective: To create a search tool that allows users to query past NFL games for detailed insights such as scores, total yards, and other metrics

Target Audience: NFL fans, journalists, and fantasy sports players seeking detailed

game statistics



Data Collection and Sources

- FiveThirtyEight NFL ELO dataset for team ratings and historical outcomes
- Pro Football Reference for game details like total yards and turnovers
- ESPN API for current game data and unstructured text commentary





Algorithm and Approach

Core Algorithm: BM25 for ranking games by relevance to the query

BM25 - Calculates scores based on term frequency and inverse document frequency, considering factors like team names, game stats, and keywords.

Example: Defining "Excitement" in Games

- High scores and close margins
- Significant game events (turnovers, high yardage)
- Fan engagement metrics or key words indicating excitement in game descriptions

Implementation Strategy

Data Processing:

Pipeline Creation: Scripts clean, standardize, and store data from APIs

Text Cleaning: Removes irrelevant characters, HTML tags, and inconsistencies in data

BM25 Scoring and Query Matching:

Assign weights to different fields

Example: Yardage and Big Plays - Metrics like total yards or big plays are given higher weights for terms like "high-scoring" or "offensive performance"

Tests to Ensures Query Results Meet Expectations

Relevance Testing: Run queries against BM25 to check top ranked games match expected results

Test Case Setup: Define several test cases based on common and unique user queries

Unique Query Capabilities of Gridiron Guru

Flexible and Specific Query Handling:

Users can ask high specific questions like:

- "Closest games for the Arizona Cardinals"
- "Most exciting games for the Seattle Seahawks"
- "Biggest blowouts for the Houston Texans"

User Benefits:

- Allows fans to explore NFL history with more specificity and relevance than traditional search engines
- Our customized ranking and filtering allows for more tailored user results

Key Challenges and Solutions

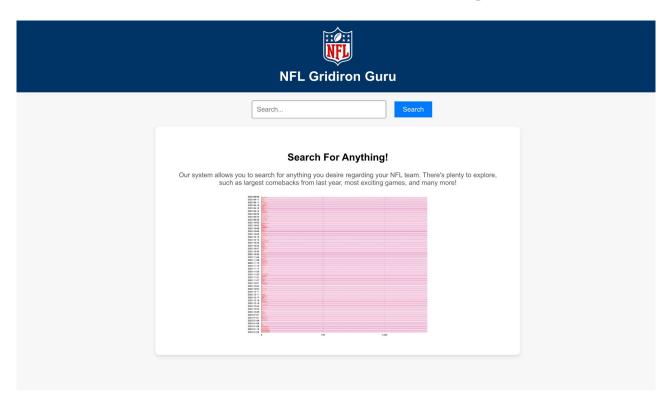
Challenges:

- Data cleaning and normalization
- Opmziting BM25 to handle large datasets efficiently
- Evaluating output quality and relevance

Solutions:

- Robust data cleaning pipelines (ie: Removing unnecessary details from data, ensuring correct formatting, and normalizing team names and player statistics across different sources)
- BM25 Fine Tuning (ie: Adjusting term frequency scaling and document length normalization and testing different weighting schemes for different fields)

Gridiron Guru Homepage Ul



Example Query Result for "most exciting cardinals games"

