**Article Name:** Stats are not enough

**Insight Name:** Confrontation Matrix Analysis for Batsman Steve Smith

**How Insights Were Extracted:**

**Feature Representation:**

* A confrontation matrix is constructed for each player using batting and bowling features extracted from deliveries they are a part of.
* Rows represent batting features, columns represent bowling features, and entries indicate the count of co-occurrence of respective feature values.

**Subset Extraction for Player-Specific Analysis:**

* A subset of text commentary is obtained based on a filter tuple with elements (player, opponent player, time, type).
* Elements:
  + Player and opponent player: Single-player or a group of players.
  + Time: Per session, per day, per innings, per match, per series, or an entire career
  + Type: Batting or bowling.

**Confrontation Matrix Construction:**

* Example: Matrix construction for batsman Steve Smith is of size 19 × 12 (rows for batting features, columns for bowling features of opponent players).
* Each element indicates how the batsman is confronted with the bowlers.

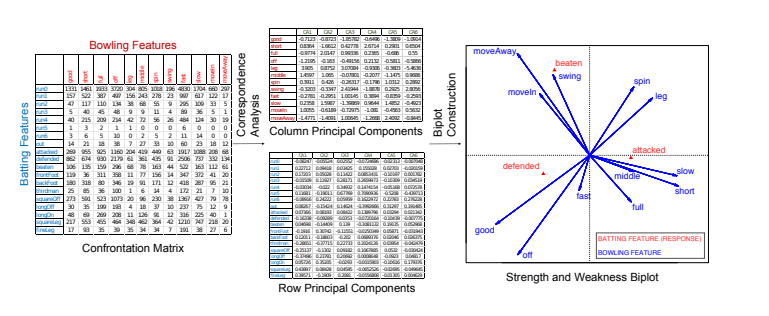
**Formulas, Limitations, and Model:**

* Formulas: The content doesn't explicitly mention formulas. The construction of the confrontation matrix involves counting occurrences of feature values.
* Limitations:
  + The method depends on the availability and accuracy of text commentary data.
  + It assumes that the chosen features effectively represent a player's performance.
* Model: The content doesn't specify a particular model but describes a systematic approach for constructing confrontation matrices for players.

**How Insights Are Visualized:**

* The constructed confrontation matrix serves as a visual representation of how the batsman, in this case, Steve Smith, confronts various bowlers.
* Each element in the matrix provides insights into specific aspects, such as how many times the batsman has shown aggression on short-length deliveries.

**Graph:**

****

**Insight Name:** Strength and Weakness Rules for Batsman Steve Smith

**How Insights Were Extracted:**

**Dimensionality Reduction:**

* Bowling and batting features are in a high-dimensional space.
* Correspondence Analysis (CA), a multivariate statistical technique, is employed to reduce dimensionality and capture discriminative variables.

**Central Idea in CA:**

* CA tests the independence of events (row variables and column variables).
* Non-independence indicates a relationship between batting and bowling features, captured in CA.

**Steps in Learning Rules:**

* Applying CA on the confrontation matrix to obtain row and column principal components.
* Plotting them on a two-dimensional biplot.

**Biplot Interpretation:**

* Row and column vectors with high inner product values indicate proximity and positive correlation.
* Strength rule: Close vectors when the selected batting feature is attacked.
* Weakness rule: Close vectors when the selected batting feature is beaten.

**Formulas, Limitations, and Model:**

* Formulas: The content doesn't explicitly provide formulas for CA but emphasizes testing the independence of events.
* Limitations:
  + The effectiveness of CA depends on the quality and representativeness of the confrontation matrix.
  + Assumes that the selected features adequately represent a player's performance.
* Model: Correspondence Analysis (CA) is the primary model for dimensionality reduction and rule extraction.

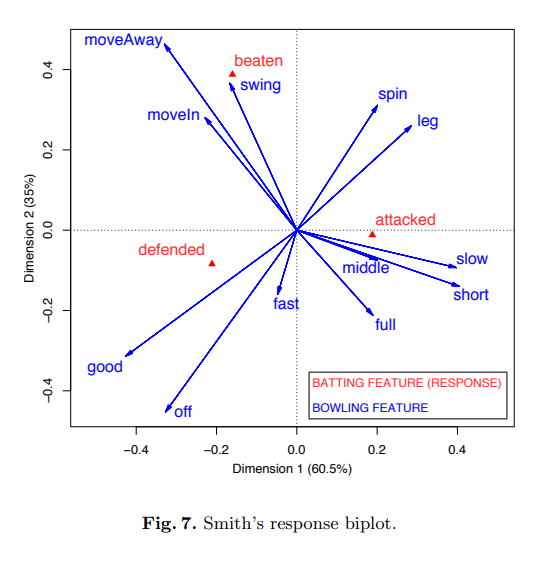
**How Insights Are Visualized:**

**Biplot Representation:**

* The two-dimensional biplot visually represents the relationship between batting and bowling features.
* Strength and weakness rules are inferred based on the proximity and correlation of vectors in the biplot.

**Visual Representation Example:**

**Biplot for Batsman Steve Smith:**

****

The inferred strength and weakness rules are derived from the proximity of vectors in the biplot.

**Insight Name:** Year-wise Changes in Strength and Weakness Rules for Batsman Steve Smith

**How Insights Were Extracted:**

**Three-Dimensional Confrontation Tensor:**

* Objective: Graphically display associations between batting features, bowling features, and time.
* Year-wise confrontation matrices for each batsman and bowler result in a three-dimensional confrontation tensor.

**Analysis Approach:**

* Three-Way Correspondence Analysis (TWCA) is employed to analyze the association between batting features, bowling features, and time.
* Temporal changes in strength and weakness rules are computed through TWCA.

**Visual Representation:**

* Inner product of principal components (batting and bowling-time) is used to reconstruct the original three-way confrontation tensor.
* Line plot visualization of inner product values represents year-wise changes.

**Formulas, Limitations, and Model:**

* Formulas: The content doesn't explicitly provide formulas for TWCA but emphasizes the inner product for numerical assessment.
* Limitations:
  + Effectiveness relies on the quality and representativeness of confrontation matrices and temporal data.
  + Assumes that selected features adequately represent a player's performance evolution.
* Model: Three-Way Correspondence Analysis (TWCA) is the primary model for analyzing three-way associations.

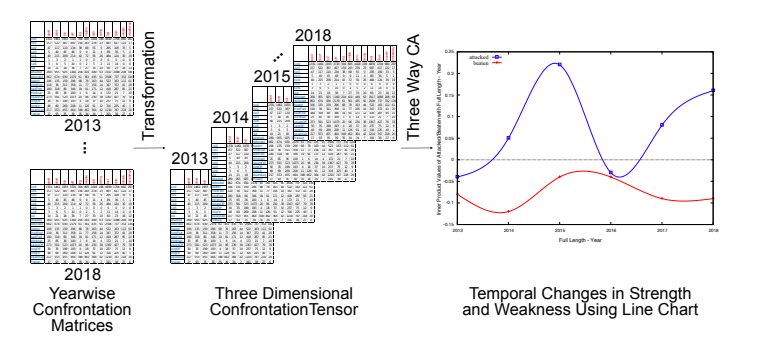
**How Insights Are Visualized:**

**Line Plot Representation:**

* Inner product values between batting features and coded bowling-time features are plotted in a line graph.
* Blue-colored line represents the strength rule (attack strategy on full-length deliveries).
* Red-colored line represents the weakness rule (beaten on full-length deliveries).

**Visual Representation Example:**

**Line Plot for Batsman Steve Smith:**



Year-wise variations indicate shifts in Smith's performance strategy.

**Potential Approaches for X-axis and Y-axis Determination:**

X-axis: Different Batting Features (e.g., aggression, defensive play, shot types)

Y-axis: Corresponding Bowling Features (e.g., pace, swing, length)

Z-axis (Time): Years or Matches

**Visualization Method:**

The mention of Correspondence Analysis (CA) and Three-Way Correspondence Analysis (TWCA) indicates that the analysis involves multivariate statistical techniques. These techniques might consider various variables, and the axes represent the principal components obtained through these analyses**.**

**Paper 2**

**Insight:** "Batsman's Shot Preference in Varying Bowling Lengths"

**Abstract**: The presented work utilizes a personalized deep neural network approach, integrating cricket-specific scenarios, such as the 2019 Cricket World Cup Final, to predict where a specific batsman is likely to hit a ball delivered by a particular bowler in a defined game situation. The model's predictive capabilities are showcased by varying the bowling length, illustrated with the example of Ben Stokes' shot predictions. The paper visualizes the predicted shot locations for different bowling lengths, demonstrating Stokes' preference for the mid-wicket region and highlighting significant magnitude changes in predictions for yorker, full toss, and short-pitched deliveries.

**Extraction Method:**

The insight is extracted through a personalized deep neural network approach, predicting the probabilities of a specific batsman's shot placement against a particular bowler and bowl type in a given game scenario. The scenario chosen for illustration is the 2019 Cricket World Cup Final between England and New Zealand, specifically focusing on a critical delivery where England needed 9 runs from 3 balls.

**Methods:**

* Deep Neural Network (DNN): The core methodology involves the application of a personalized deep neural network.
* Scenario Integration: Specific game scenarios, like the 2019 Cricket World Cup Final, are incorporated to enhance the model's contextual relevance.
* Visualization Technique: The bowling length is systematically varied while keeping other trajectory aspects fixed to visualize the predicted shot likelihoods.
* Magnitude Changes: Insights are drawn by observing magnitude changes in predicted shot locations for different delivery lengths.

**Applications:**

* Team Performance Strategies: The model provides a foundation for teams to develop pre-game strategies and dynamic in-game tactics based on evolving match contexts.
* Media Storytelling: Media outlets can leverage the model's predictions to craft richer and more engaging storylines, moving beyond conventional score and win predictors.

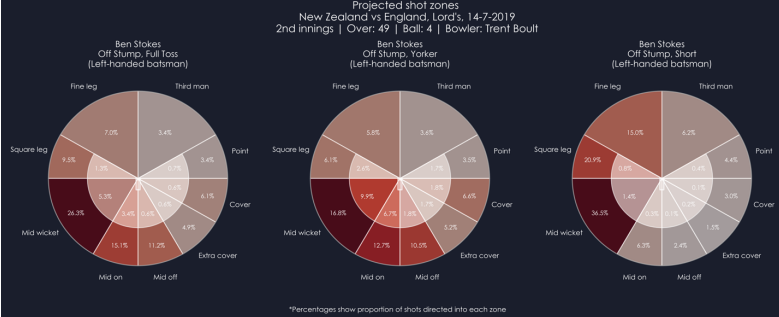
**Limitations:**

* While the model provides valuable insights, its accuracy may be influenced by unforeseen factors during a live match.
* The predictions are based on historical data and may not fully capture a player's real-time form and strategy changes.

**Formulas:**

* The model uses intricate neural network architectures to compute the probabilities. The variation in predicted shot likelihoods concerning different bowling lengths is calculated by modifying the length variable while keeping other trajectory aspects fixed.

**Visualization:**

Below figure displays the visualization of the predicted shot likelihoods for Ben Stokes in response to a delivery with varying lengths. The outfield zone in the mid-wicket region is identified as Stokes' preference for the given delivery line. The magnitude of the predicted likelihood undergoes discernible changes—approximately 10% and 20%—when transitioning from a yorker to a full toss and a short length, respectively.

**Insight Name:** "Contextual Shot Prediction in Cricket"

**Insights Extraction:**

The content focuses on the challenging nature of predicting a batsman's shot type and direction in cricket, emphasizing the specificity of each situation based on contextual features such as the identity of the batsman, bowler, fielders, and the game state. The limitation of current methods, which often rely on broad averages without considering contextual nuances, is highlighted.

**Methods:**

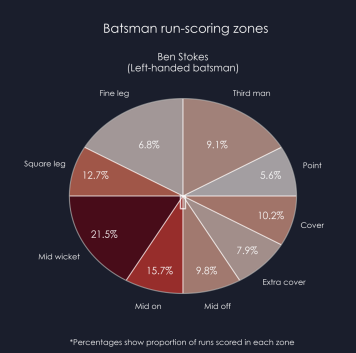
* Contextual Feature Analysis: Develop algorithms to analyze contextual features such as player identity, ball trajectory, bowler, and match situation. This involves assigning weights to each feature based on its relevance and impact on shot prediction.
* Player-specific Models: Create personalized models for each batsman, considering their strengths, weaknesses, and shot preferences. Machine learning models, such as deep neural networks, can be trained on historical data to predict a batsman's likely response to different deliveries.
* Trajectory and Context-Aware Analytics: Introduce trajectory-based analysis that considers the path of the ball and its potential impact on shot selection. Context-aware analytics should adapt to specific match situations, dynamically adjusting predictions based on real-time game states.

**Limitations:**

* The complexity of predicting shots in cricket necessitates sophisticated models, and there may be inherent uncertainties due to the unpredictability of player behavior.
* The accuracy of predictions could be affected by rapidly changing match scenarios and the introduction of new playing styles.

**Visualization Methods:**

* Contextual Heatmaps: Visualize shot prediction probabilities using heatmaps that incorporate contextual features. Each heatmap could represent the likelihood of a shot landing in different areas of the field based on the specific context.
* Dynamic Trajectory Charts: Create dynamic charts that illustrate how a batsman's shot preferences change based on ball trajectory, bowler identity, and other contextual factors. This could involve interactive charts that adjust in real-time during a match.

****

**Insight Name:** "Personalized Shot Type Prediction in Cricket"

**Insights Extraction:**

The content describes an innovative approach to revolutionize the prediction of matchups between bowlers and batsmen in cricket. The primary goal is to define shot types based on the intent and direction of the batsman rather than relying solely on shots end locations. The proposed method involves personalized deep learning to predict the likelihood of specific shot types for a given delivery.

**Methods:**

**Intent-Based Shot Analysis:**

* Definition of Shot Types: Shift focus from shot end locations to shot intent and direction.
* Aggression Labels: Assign labels (0, 1, 2) indicating increasing aggression based on exploratory analysis of shot labels and corresponding runs scored.

**Data and Model:**

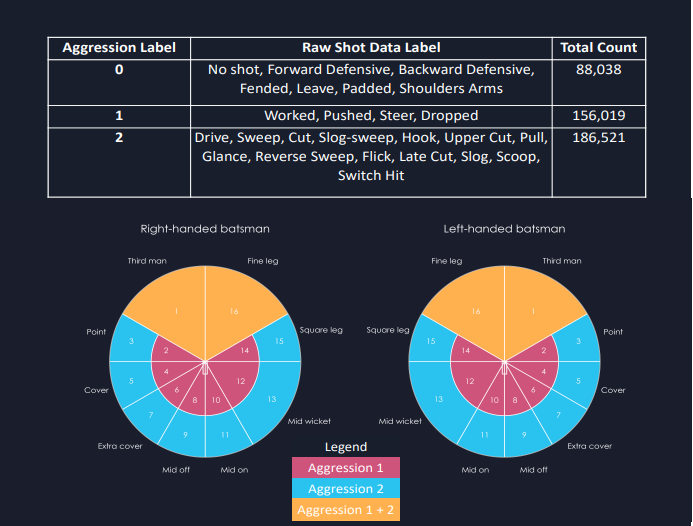
* Dataset: Utilize Opta's ball-by-ball data from the past 8 years of international cricket (over 430,000 balls).
* Deep Learning Model:
  + LSTM for Ball-by-Ball Data: Use a multi-layered long short-term memory (LSTM) recurrent neural network for ball-by-ball data, incorporating match context and delivery trajectories.
  + Feed-Forward Neural Network: Augment the LSTM with a multi-layered feed-forward neural network containing personalized information for the batsman and bowler of each delivery, based on their historical data.

**Limitations:**

* The effectiveness of the model may be influenced by the dynamic and unpredictable nature of cricket gameplay.
* The quality of predictions could be impacted by variations in player form and strategy changes over time.

**Visualization Methods:**

* **Shot Type Heatmaps:**
  + Target Variables: Create bespoke target variables based on shot angles and aggression labels, splitting the field into 16 zones.
  + Visualization: Display shot type likelihoods using heatmaps that represent the intent and shot angles rather than shot end locations.
* **Personalized Insights Dashboards:**
  + Batsman and Bowler Profiles: Develop individualized dashboards for batsmen and bowlers, providing insights into their shot preferences and historical performances.
  + Interactive Elements: Include interactive elements allowing users to explore personalized insights based on specific players, match contexts, and shot types.
* **Trajectory Analysis Charts:**
* Visualizing Delivery Trajectories: Incorporate charts that visually represent the trajectories of different deliveries, enhancing the understanding of how shot types are influenced by ball paths.



**Insight Name:** "Shot Type Prediction in Cricket: Integrating Delivery Trajectory and Match Context"

**Extracted Insights:**

The study aims to revolutionize predictions of cricket matchups between bowlers and batsmen by introducing a personalized deep learning approach. Focusing on shot type prediction, the research incorporates detailed ball-by-ball delivery information, including line, length, ball movement, bowler characteristics, and delivery angles. To enhance context, match-specific features such as stage of the innings, wickets taken, and runs scored are integrated, providing a comprehensive understanding of the factors influencing shot selection. The model's development involves memoryless models like Random Forest and Feed-Forward Neural Network, with significant performance gains demonstrated. Notably, the study utilizes a multi-layered LSTM for time series analysis, considering the sequential nature of cricket matches. This LSTM model achieves a remarkable accuracy of 21.0% and a log loss of 2.45. The insights are visualized through shot probability heatmaps, dynamic shot type graphs, and feature importance charts, providing a comprehensive and interactive understanding of shot type predictions in various cricketing scenarios. The limitations acknowledge the inherent complexity of cricket gameplay and the potential variability in model performance based on player-specific behaviors.

**Methods:**

1. **Feature Integration:**

* Delivery Information
  + Line and length of the ball.
  + Movement of the ball (through the air and off the pitch).
  + Handedness and style of the bowler.
  + Bowling angle (over the wicket or around the wicket).
* Match Context Features:
  + Stage of the innings.
  + Wickets taken by the bowling team.
  + Runs scored by the batting team.
  + Batsman-specific features (current runs scored, deliveries faced).

1. **Modeling Approaches:**

* Memoryless Models:
  + Random Forest: Achieved 19.9% accuracy and 2.51 log loss.
  + Feed-Forward Neural Network: Improved accuracy to 20.3% and log loss to 2.48.
* Time Series Analysis:
  + Multi-Layered LSTM: Utilized a lookback window of 6 (matches deliveries per over), resulting in 21.0% accuracy and 2.45 log loss.

**Limitations:**

* The model's accuracy may be influenced by the dynamic and unpredictable nature of cricket gameplay.
* Decision-making factors for shot selection are complex and may not be fully captured by the provided features.
* Model performance could vary based on player-specific behaviors and strategies.

**Visualization Methods:**

1. **Shot Probability Heatmaps:**

* Visualization of Shot Types: Create heatmaps representing the probability distribution of shot types across the 17 zones.
* Incorporate Context Features: Enhance heatmaps by incorporating context features to visualize how shot preferences change based on match situations.

1. **Dynamic Shot Type Graphs:**

* Time Series Visualization: Develop dynamic graphs illustrating how shot types evolve over time during a match, considering both delivery information and match context.
* Interactive Elements: Include interactive elements for users to explore shot type trends in specific match situations, innings stages, or player contexts.

1. **Model Performance Metrics:**

* Accuracy and Log Loss Plots: Visualize the performance metrics (accuracy and log loss) over training epochs for each modeling approach.
* Comparison Charts: Develop charts comparing the performance of memoryless models (Random Forest, Feed-Forward Neural Network) with the LSTM model.

1. **Contribution of Features:**

* Feature Importance Charts: Illustrate the contribution of different features to the prediction of shot types.
* Contextual Influence: Highlight features that significantly influence shot selection, considering both delivery and match context.