Analysis of the Effect of High Dew Points on Baseball Pitches

Objective:

Determine the influence of high dew points (indicative of high humidity) on each pitch in a baseball game.

Procedure:

1. **Exploratory Data Analysis (EDA):**

   - Checked for missing values in the dataset.

   - Visualized the distribution of features potentially affected by humidity such as vertical and horizontal breaks, spin rate, release speed, and approach angles.

   - Identified features with varying distributions, some showing central tendencies while others were more dispersed.

2. **Feature Engineering:**

   - Created interaction features to possibly capture more complex relationships between attributes affected by humidity.

   - Features such as spin_speed_interaction, spin_vbreak_interaction, and others were derived from existing attributes.

3. **Model Building:**

   - Proposed the use of the Isolation Forest algorithm for anomaly detection. This model treats pitches affected by high dew points as "anomalies."

   - The anomaly score from this model can be interpreted as the probability that a pitch was affected by a high dew point.

4. **Evaluation:**

   - Visualized the distribution of computed probabilities to understand the likelihood of each pitch being affected by high dew points.

   - Observed that many pitches had low probabilities, while a subset exhibited higher chances of being influenced by high humidity.

Conclusion:

The project provides a structured approach to infer the effect of high dew points on baseball pitches using data-driven techniques. The results, based on the

patterns observed in the data and the assumptions of the Isolation Forest model, offer insights into the potential influence of humidity on pitch behaviour.