Project Proposal: Clustering and Dimension Reduction in Pokémon Data to Optimize Player Strategy

Jin Huang 404899 Georgii Bykov 444449 Jakub Gazda 419272

Introduction

The Pokémon franchise features a variety of creatures with different attributes, making it a rich dataset for analysis. This project aims to employ clustering and dimension reduction techniques to identify Pokémon types with higher attack values, assisting new players in focusing their efforts on catching these powerful Pokémon. Additionally, we will explore which attributes are most crucial in defining a Pokémon's characteristics.

Objectives

Clustering Analysis: Implement and compare K-means, PAM (Partitioning Around Medoids), and CLARA (Clustering Large Applications) algorithms to cluster Pokémon based on their attributes, with a focus on attack values.

Identify High-Attack Pokémon Clusters: Analyze the clusters formed by each algorithm to identify those with the highest average attack values and determine the common Pokémon types within these clusters.

Dimension Reduction Analysis: Utilize Singular Value Decomposition (SVD) and Principal Component Analysis (PCA) to identify the most significant attributes that define a Pokémon's characteristics.

Methodology

Data Collection and Preprocessing:

Collect data on Pokémon attributes including type, attack, defense, speed, and special abilities.

Clean and preprocess the data, handling missing values and standardizing numerical attributes for uniform analysis.

Clustering Analysis:

Implement K-means, PAM, and CLARA algorithms on the preprocessed data.

Evaluate the clustering results using metrics such as silhouette score and within-cluster sum of squares to determine the optimal clustering method.

Cluster Evaluation and Type Identification:

Analyze the clusters formed by each algorithm to identify those with the highest attack values.

Examine the Pokémon types prevalent in these high-attack clusters to recommend to new players.

Dimension Reduction:

Perform SVD and PCA on the dataset to identify the key attributes that contribute to a Pokémon's power and uniqueness. Analyze the loadings of the principal components to determine which attributes are most influential in defining a Pokémon.

Expected Outcomes

A clear comparison of K-means, PAM, and CLARA clustering methods in the context of Pokémon data, highlighting the most effective approach for identifying high-attack Pokémon.

A list of Pokémon types that frequently appear in high-attack clusters, providing new players with strategic insights on which Pokémon to target.

A comprehensive analysis of the most important Pokémon attributes, aiding in the understanding of what makes a Pokémon powerful and distinctive.

Significance

This project will not only help new Pokémon players by informing their catching strategies but also contribute to the field of data science by applying clustering and dimension reduction techniques to a popular and engaging dataset. By understanding the attributes that define powerful Pokémon, players and game designers alike can gain valuable insights into game dynamics and strategy optimization.