

Identifying an optimal feature set to analyse species richness in Butterflies

Erin Aho, Albert Nyarko-Agyei, and Soham Talukdar

Butterflies are important taxa, both ecologically and culturally, but have been declining in population despite conservation efforts. To improve understanding of where to focus conservation efforts we aim to identify which features are the most informative in terms of species richness in an area. The species count of butterflies in 45 broad regions was analysed. Based on existing research, we decided on a set of 29 features, both bioclimatic and anthropogenic, which are known to affect species richness. We generated a dataset by taking the mean across our regions for each feature, using open-source GIS tools. After researching current feature selection methods, we decided to use Joint Mutual Information (JMI) as our selection criterion, due to the high amount of mutual information in many of our features, and the strong performance JMI has been shown to have form small data samples. After performing feature selection on our dataset using JMI, we were able to identify that the most informative set of 5 features is Annual Precipitation, Isothermality, Mean Diurnal Temperature Range, Precipitation of the Warmest Quarter, and the Percent of Urbanization. We discuss possible limitations of our approach, and areas further research is needed in order to better understand butterfly species richness.

Introduction.—

Method.—

Results.—

Discussion.—

Conclusion.—