Classification of cyclist types

Master‘s Thesis of Walter Garcia

Mentoring:
M.Sc. Victoria Dahmen
Dr.-Ing. Anna Takayasu

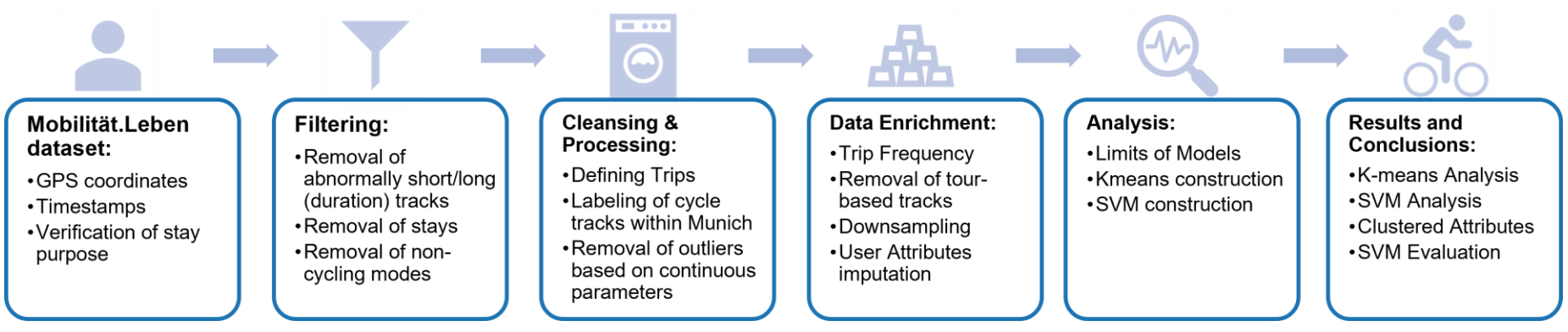

Figure 1 - Procedure for Classification Development

In response to the growing popularity of cycling as a mode of transportation, this thesis delves into the imperative task of comprehending cycling behavior, recognizing the multifaceted influences on route choices, including individual preferences and the initial mode selection. Departing from the prevalent reliance on stated preference, this research leverages revealed preference data obtained through GPS trajectories from the Mobilität.Leben project.

We begin by employing two distinct machine learning methods, an unsupervised model called, the k-means clustering model to discern various driving characteristics and then move forward with a supervised classification model called Support Vector Machines (SVM) model to discern the most effective binary classification based on cyclists engaged in commuting, recreational, home-based, and errand-related trips. It is important to acknowledge that the quality of results obtained from machine learning models is linked to the quality of the underlying data and will elaborate on the variables used.

The thesis used continuous variables relevant to cycling classification and set aside categorical variables for observational purposes. The Mobilität.Leben project had over 25 continuous and categorical variables available but, after literature revealed 7 continuous variables be optimal. The decision to use continuous variables was a preventative measure to binary attributes that can arise by using categorical variables in spatial models like Kmeans and SVM models that rely on Euclidean distances (Harrington, 2012c; Huang, 1998). The variables used are age (yr.), trip start time (hr.), trip duration (hr.), trip average speed (kph), trip length (km), trip cycling facilities use (%), and node count (1/km) processed into a higher dimension and embodied in a spatial format to represent cyclists' conditions and later be evaluated by both models.

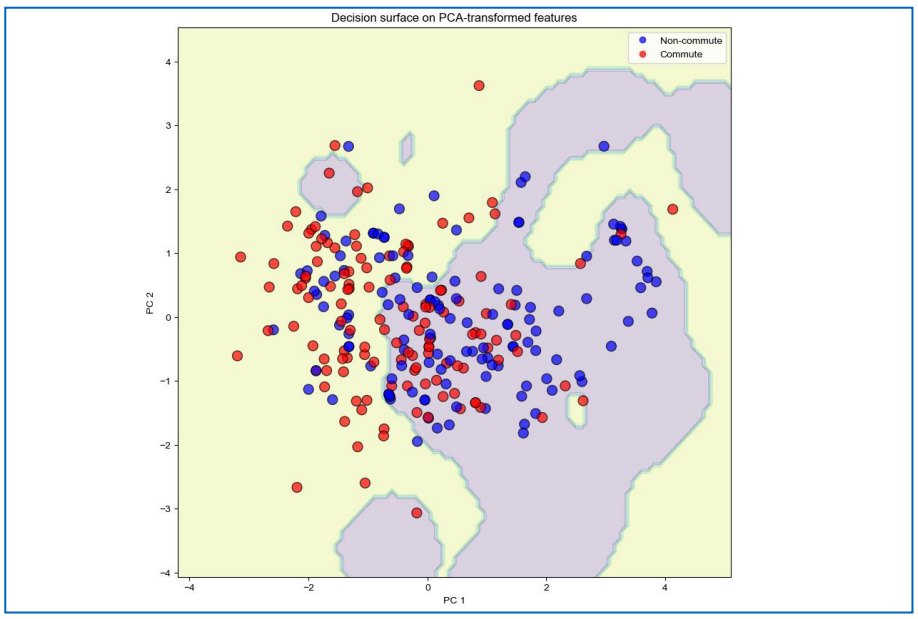

Figure 3 - Optimized SVM Hyperplane: Commute trips

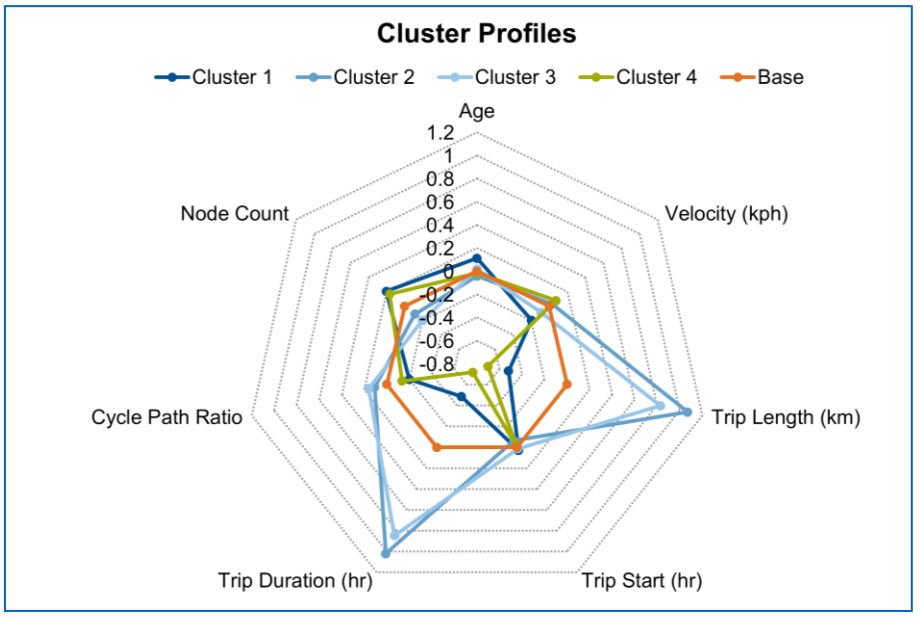

Figure 2 – Radar Graph Cluster profiles (Kmeans)

The results demonstrate the Kmeans exploratory analysis having commonalities among cycling trip characteristics Cluster 4 being the more home trip dominant cluster, Cluster 3 and 2 being more errands trip dominant and Cluster 1 commute dominant regarding cycling trip share. As for the radial basis function (RBF) based SVM model the findings reveal being the most accurate in predicting cycling commute trips and with minimal optimizing of the shape and regularization parameters of the radial basis function based on several versions generated via OneVsAll binary model philosophy. To be exact the SVM model showed an accuracy of 85% when in came to predicting commute trips. To conclude, the Kmeans cluster analysis approach brings exploratory insight to variable selection for the continuous and categorical variables and compliments the SVM model in optimally choosing continuous variables for the predictive capabilities and bringing insight into trip-based cycling classification.

Access to the full thesis and references here: