The training data are $\{B_n, y_n\}$, $n \in [1, N]$, N is the total number of day_geo bins, $B_n = \{b_{nm}\}, m \in [1, M]$, M is a constant as the number of selected user for each day_geo bin. $B_{nm} = \{b_{nmk}\}, k \in [1, K]$, K is a constant as the number of images selected for each user. $B_n \subset \{U_n\}$ and $B_{nm} \subset \{I_{Un}\}$. For example, each $U_n = \{u_i\}, i \in [1, N_{Un}]$ represents all the users in each day_geo bin and $Img_{nm} = \{Img_j\}, j \in [1, N_{UnImg}]$ are all the images from each user. The descriptor set is D for each entry in set B.

Our objective function is $f : B \rightarrow Y$. $d_{nmk} = \phi_I(\rho_I(Img_{nm}))$ and $d_{nm} = \phi_U(\rho_U(U_n))$. Function ϕ is to aggregate the features of a constant number of entries selected by function ρ. Finally we'll learn a MLP to get $d_n = \Phi(d_{nm})$.