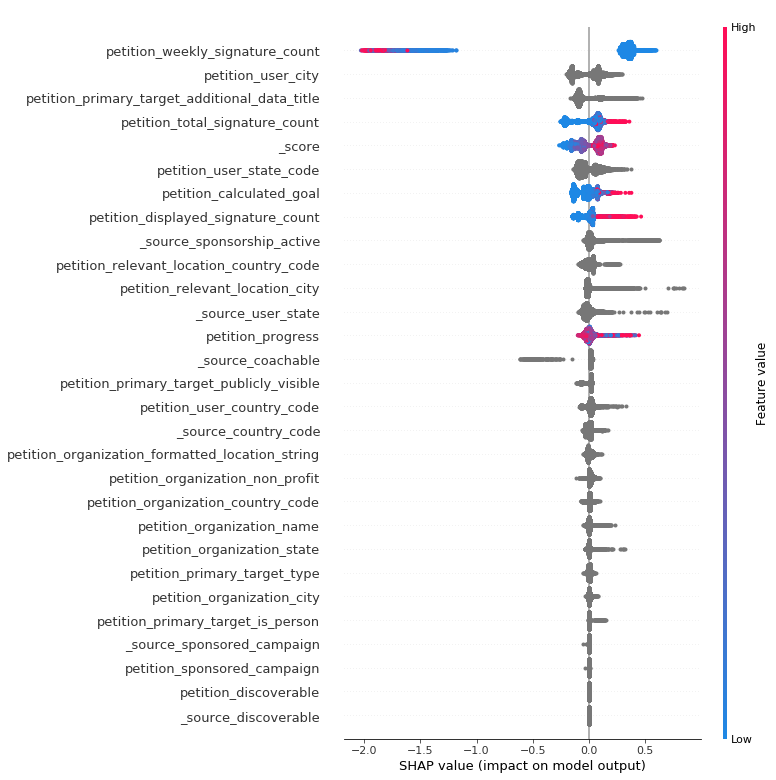
Link to visualizations- <https://public.tableau.com/profile/varun3553#!/vizhome/PDC_Hackathon_2019/PDCHackathon2019>

* Derived correlation of features such as weekly signature count, score, petition target and state using **shap**
* Analysed success rate and number of petitions over geographies



Feature Analysis-

* Carefully combed through all available feature to understand their complexity and integrity
* Removed highlight columns for victory prediction as they were disjoint from source and petition columns
* Classified columns into feature types such as categorical, numeric, etc.



Feature Preparation-

* HTML tag extraction using Regex in Python
* Dropped petition\_petition\_status as it was identified as target leak

Model Selection-

* Used catboost classifier
* Catboost classifier allows us to use categorical columns without explicit one-hot encoding
* 80:20 split of train data into X\_train and X\_test
* 70:30 split of X\_train into X\_train and X\_eval
* Similar split for target data
* Test accuracy - 90.85% for victory prediction and 100% for category classifier
* Eval accuracy - 90.74% for victory prediction and 100% for category classifier

Parameter Tuning-

* Hyperparameter tuning using sklearn.randomsearchcv()
* Tuned on parameter: learning\_rate, depth, l2\_leaf\_reg