**Overview**

Attached are 2 data sets consisting of matching scores between personal details of a person to records of individual contributions to political parties in the USA (see attached screenshot for example data).

Each attribute in the data sets attached represents a matching score between personal data to specific contribution field. e.g " 200~Jaro-Winkler" represents the Jaro-Winkler string distance (based on Levenshtein distance) between matched person's name and "Contributor Name" field in contribution records. Higher scores represent better matches. BEWARE: some of the attributes are irrelevant to the “label”.

**Study task**

Your task is to create a predictive model that will predict the "label" attribute in the TestDataset (based of course on the model you trained on TrainDataset). Please provide a table with RowId and your PredictedLabel. You can do it using any tool you favor, but please submit your code.

Once you’ve completed the model we’d like to see one (or two) page summary that includes

·         a description of the analysis of the data.

·         the methods used to predict the “label”

·         difficulties / limitation of the given dataset

·         suggestions for further studies

We don’t expect you to reach specific accuracy level or use specific ML methods.

We rather focus on how you deal with the given task, the statistical and mathematical tools you choose and mainly on how you reason your choices.

Best of luck,

Intelligo-group