Doplnit chybajuce v:

* news\_front\_page – dam ze default – 0 – pretoze vecsinou stranka nie je titulka novin,, takisto sa ale moze pouzit na doratanie nejaky klasifikator
* Alchemy category a score – doplnit chybajuce (mozno klasifikator vlastny natrenovat)

Outliere v:

image\_ratio

Treba spravit funkciu ktora bude odjebavat outliere od nejakeho rozdahu a doplnat ich za priemernu hodnotu,, argumenty funkcie budu rozsah

* Normalizacia zavisi od vyberu techniky klasifikatora, ale neuskodi spravit funckiu

**Normalizacia**

import pandas

TRAIN\_DF = pandas.read\_csv('D:/train.tsv', sep='\t');

for index in range(1,TRAIN\_DF['image\_ratio'].count()):

TRAIN\_DF['image\_ratio'][index] = (TRAIN\_DF['image\_ratio'][index] - TRAIN\_DF['image\_ratio'].min())/(TRAIN\_DF['image\_ratio'].max() - TRAIN\_DF['image\_ratio'].min())

**Chybajuce v page news**

for index in range(1,TRAIN\_DF['page\_news'].count()):

if (TRAIN\_DF['page\_news']index] == “”)

TRAIN\_DF['page\_news']index] = 0.

**Outlier image ratio**

for index in range(1,TRAIN\_DF['image\_ratio'].count()):

if (pandas.to\_numeric(TRAIN\_DF['image\_ratio'][index]) < 0) or (pandas.to\_numeric(TRAIN\_DF['image\_ratio'][index]) > 1):

TRAIN\_DF['image\_ratio'][index] = TRAIN\_DF['image\_ratio'].avg()