files starting by A\_detect\_potential\_stigma\_clean refer to the sensitive data (positive score) which is extracted from dataset version 2, and is cleaned and duplicates tweets are also removed.

files starting by B\_detect\_potential\_stigma\_clean refer to the non-sensitive data (zero scores) which is extracted from dataset version 2, and is cleaned and duplicates tweets are also removed.