Aim：classify death reasons described in natural language by death reasons type

Data:

1. death reasondo.txt A txt files that record the death reasons described in natural language to be classified. Need to be modified before you turning the format into csv.

文本

描述已自动生成

1. deathreason.csv A List of standard death reasons. It is a dictionary, don’t do any change to it.
   1. icd10: standard icd10 code
   2. disease: the corresponding name of disease
   3. Inf\_sexual Inf\_resp Inf\_Enteric Inf\_tropical Inf\_other Maternal Neonatal Nutritionaldeficiencies Neoplasms cvd chronic\_resp Dis\_Digestive Dis\_Neurological Dis\_Mental Dis\_Substanceuse Diabetes Chronickidney Skin Senseorgan Glaucoma Cataract Musculoskeletal Congenitalbirthdefects Urinary\_maleinfertility Gynecological Hemoglobinopathies\_hemolytic Endocrinemetabolicimmune Dis\_oral Inj\_Transport Inj\_Unintentional: Class of standard death reasons,

Suggestion:

1. Generate empty variables (Inf\_sexual Inf\_resp Inf\_Enteric Inf\_tropical Inf\_other Maternal Neonatal Nutritionaldeficiencies Neoplasms cvd chronic\_resp Dis\_Digestive Dis\_Neurological Dis\_Mental Dis\_Substanceuse Diabetes Chronickidney Skin Senseorgan Glaucoma Cataract Musculoskeletal Congenitalbirthdefects Urinary\_maleinfertility Gynecological Hemoglobinopathies\_hemolytic Endocrinemetabolicimmune Dis\_oral Inj\_Transport Inj\_Unintentional) in the csv file you rewrite from death reasondo.txt.
2. Classify the death reasons listed in death reasondo.txt uniquely into standard death reasons as listed in deathreason.csv according to its similarity with variable [disease].