

## Wikidata - the Sherlock Holmes investigative AI

Given personal information about a person determine the probable cause of death.

Personal information is provided in wikidata format, for example: [Abraham Lincoln](#). Given all other information and entity links you need to build a classifier that would determine the most probable manner of death (you should not use 'manner of death' or 'cause of death' fields as your input data).

There are many such manners defined in wikidata, but we are interested in the 4 most common ones:

natural causes	29764
suicide	6094
accident	5384
homicide	4523

You can use the whole [wikidata dump](#), or SPARQL queries to gather the data to use. You need to devise training and evaluation procedure yourself. We are interested in a solution with good performance, that would be easy to improve and evaluate. It should also be built in a limited amount of time - a minimal working proof-of-concept will do.

See also:

[Most common high-level causes](#)

[Overall causes of death ranking](#)