Classification of Cause of death

Business Problem:

Even in well-regulated hospital service, medical officers misinterpret the cause of death. This project aims to build a classification model to find the cause of the death of an individual.

Aim:

To classify patients' cause of death

Data Description:

The data provided is for 2020 death cases. There are 16 features available with case number as a unique identifier and Reason of Death as a target variable. The Data dictionary is attached with the data.

RedShift Database:

The data is present in Redshift within three different tables with a total of 47406 data points. The description of the tables is as follows:

1. personal

   It contains patient personal data with the case number, date, age years, age months, sex, insurance group, number of family members, and country of origin as features.

2. accident

   It contains details related to the accident such as case number, reported by, place of incident, fire, attendees group number, and patient verbal statement.

3. hospital

   This is the information provided by the hospital. It contains case number, total visit counts, place of the death, diagnosis code 1, diagnosis code 2, diagnosis code 3, p1, p2, p3, alcohol, drug, and reported reason of death features.

4. data_dictionary

   This table provides information on the features that should be considered in the analysis.

   Note:
   1. **The data should be fetched only for the year 2020.**
   2. **The columns to be considered are present in the data dictionary.**