**Abstract**

The data is dedicated to the classification problem related to the post-operative life expectancy in the lung cancer patients: class 1 - death within one year after surgery, class 2 - survival.

**Data Set Information**

The data was collected retrospectively at Wroclaw Thoracic Surgery Centre for patients who underwent major lung resections for primary lung cancer in the years 2007-2011. The Centre is associated with the Department of Thoracic Surgery of the Medical University of Wroclaw and Lower-Silesian Centre for Pulmonary Diseases, Poland, while the research database constitutes a part of the National Lung Cancer Registry, administered by the Institute of Tuberculosis and Pulmonary Diseases in Warsaw, Poland.

The data folder contains the original data file in CSV format, converted from Weka ARFF. The *Data\_Wrangling* Jupyter notebook file shows how I changed the original data set into the one used for this project.

|  |  |
| --- | --- |
| **Attribute** | **Description** |
| **Diagnosis** | ICD-10 codes for primary and secondary as well multiple tumors if any |
| **FVC** | Amount of air which can be forcibly exhaled from the lungs after taking the deepest breath possible |
| **FEV1** | Volume that has been exhaled at the end of the first second of forced expiration |
| **Performance** | Performance status on Zubrod scale, Good (0) to Poor (2) |
| **Pain** | Pain before surgery (T = 1, F = 0) |
| **Haemoptysis** | Coughing up blood, before surgery (T = 1, F = 0) |
| **Dyspnoea** | Difficulty or labored breathing, before surgery (T = 1, F = 0) |
| **Cough** | Symptoms of Coughing, before surgery (T = 1, F = 0) |
| **Weakness** | Weakness, before surgery (T = 1, F = 0) |
| **Tumor\_Size** | T in clinical TNM - size of the original tumor, 1 (smallest) to 4 (largest) |
| **MI\_6mo** | Myocardial infarction (Heart Attack), up to 6 months prior (T = 1, F = 0) |
| **PAD** | Peripheral arterial diseases (T = 1, F = 0) |
| **Smoking** | Patient smoked (T = 1, F = 0) |
| **Asthma** | Patient has asthma (T = 1, F = 0) |
| **Age** | Age at surgery |
| **Death\_1yr** | 1 year survival period - (T) value if died (T = 1, F = 0) |
| **Diabetes\_Mellitus** | Type 2 diabetes mellitus (T = 1, F = 0) |