Cancer Type Data

Francesco Lescai

2023-06-08

Dataset

The dataset includes the following data:

tumour_type	age	staining_01	staining_02	staining_03
not_tumour	68.13368	0.2123266	0.4163561	0.5770442
unclassified_malignant_tumour	70.83473	0.2035516	0.2584667	0.7606623
adenocarcinoma	71.63894	0.1741881	0.7062178	0.6393702
not_tumour	68.44474	0.5497150	0.3540431	0.8587699
not_tumour	68.95009	0.2308089	0.4995580	0.3675874
unclassified_malignant_tumour	67.94134	0.5946114	0.4684277	0.6887187

staining_04	cell_size	cell_shape	growth_rate	presence_metastasis
0.2145811	6	oval	0.1132941	yes
0.1634798	4	spindle	0.7063195	yes
0.1036835	5	star	0.4732406	yes
0.1815295	3	round	0.1127585	yes
0.0459623	1	spindle	0.0496342	yes
0.2490539	5	spindle	0.3039531	yes

- tumour_type: diagnostic classification of the tumour
- age: age of the patient
- staining_01: histological staining used
- staining_02: histological staining used
- staining_03: histological staining used
- staining_04: histological staining used
- cell_size: dimension of the cell in microns
- cell_shape: shape of the cell
- growth_rate: growth rate of the tumour in percentage compared to previous measurement
- presence_metastasis: presence or not of metastatic cells

Assignment

Please analyse this dataset using the most appropriate methods. Prepare a report discussing your choices step by step, and presenting a data-driven justification for the analytical decisions you made.

Provide evidence, if appropriate, of relationships of dependencies in the dataset, explaining how some of the variables might influence your findings.

Discuss in the report, where appropriate, any biological background which might support your findings.

Use the most appropriate computing environment to carry out this work, and explain the code and the choice you made in a dedicated section of the report.