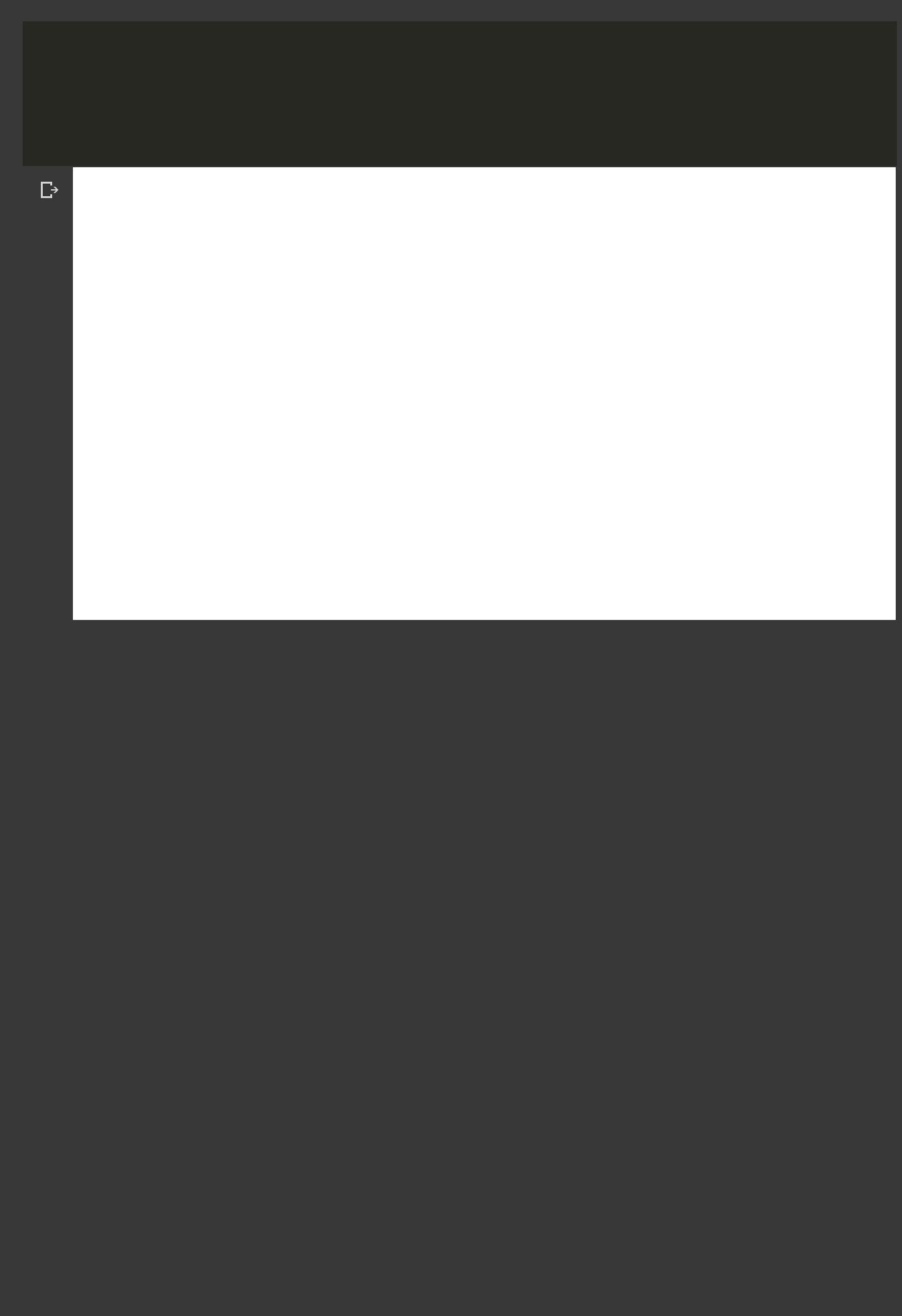
9/19/21, 10:41 PM Lab 05 - CSE422.ipynb - Colaboratory



* import pandas as pd

2 import numpy as np

3

4 dataset = pd.read\_csv('/content/mushroom edibility classification dataset.csv') 5 dataset.head()

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  | **stalk-** |  |
|  | **Unnamed:** | **class** | **cap-** | **cap-** | **cap-** | **bruises** | **odor** | **stalk- stalk- surface-** | | |  |
|  | **0** |  | **shape** | **surface** | **color** |  |  | **shape** | **root** | **above-** |  |
|  |  |  |  |  |  |  |  |  |  | **ring** |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| **0** | 0 | poisonous | 5.0 | 2 | 4.0 | bruises | 6 | 0 | 3 | 2 |  |
| is there |  |
|  |  |  |  |  |  |  |  |  |  |  |
| **1** | 1 | edible | 5.0 | 2 | 9.0 | bruises | 0 | 0 | 2 | 2 |  |
| is there |  |
|  |  |  |  |  |  |  |  |  |  |  |
| **2** | 2 | edible | 0.0 | 2 | 8.0 | bruises | 3 | 0 | 2 | 2 |  |
| is there |  |
|  |  |  |  |  |  |  |  |  |  |  |
| **3** | 3 | poisonous | 5.0 | 3 | 8.0 | bruises | 6 | 0 | 3 | 2 |  |
| is there |  |
|  |  |  |  |  |  |  |  |  |  |  |
| **4** | 4 | edible | 5.0 | 2 | 3.0 | no | 5 | 1 | 3 | 2 |  |
| bruises |  |
|  |  |  |  |  |  |  |  |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| 1 | dataset.shape | |  |
|  |  |  |  |
|  | (3124, | 20) |  |
|  |  | |  |
|  |  | |  |
| 1 | dataset.isnull().sum() | |  |
|  |  | |  |
|  | Unnamed: 0 | | 0 |
|  | class |  | 0 |
|  | cap-shape | | 8 |
|  | cap-surface | | 0 |
|  | cap-color | | 8 |
|  | bruises |  | 0 |
|  | odor |  | 0 |
|  | stalk-shape | | 0 |
|  | stalk-root | | 0 |
|  | stalk-surface-above-ring | | 0 |
|  | stalk-surface-below-ring | | 0 |
|  | stalk-color-above-ring | | 0 |
|  | stalk-color-below-ring | | 0 |
|  | veil-type | | 0 |
|  | veil-color | | 0 |
|  | ring-number | | 0 |
|  | ring-type | | 0 |
|  | spore-print-color | | 0 |
|  | population | | 0 |
|  | habitat |  | 0 |
|  | dtype: int64 | |  |
|  |  | | |
|  |  | | |
| 1 | from sklearn.impute import SimpleImputer | | |

https://colab.research.google.com/drive/1aVPQaYSVGyHYRzai0uhXgqrrE6YPTg4A#scrollTo=LmhvPZuU7wu1&printMode=true 1/4

9/19/21, 10:41 PM Lab 05 - CSE422.ipynb - Colaboratory

* impute = SimpleImputer(missing\_values=np.nan, strategy='mean')

3 impute.fit(dataset[['cap-shape']])

4 dataset[['cap-shape']] = impute.transform(dataset[['cap-shape']])

5 impute.fit(dataset[['cap-color']])

6 dataset[['cap-color']] = impute.transform(dataset[['cap-color']])



* dataset.info()

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| <class 'pandas.core.frame.DataFrame'> | | | |  |
| RangeIndex: 3124 entries, 0 to 3123 | | |  |  |
| Data | columns (total 20 columns): | |  |  |
| # | Column | Non-Null Count | | Dtype |
| --- | ------ | -------------- | | ----- |
| 0 | Unnamed: 0 | 3124 | non-null | int64 |
| 1 | class | 3124 | non-null | object |
| 2 | cap-shape | 3124 | non-null | float64 |
| 3 | cap-surface | 3124 | non-null | int64 |
| 4 | cap-color | 3124 | non-null | float64 |
| 5 | bruises | 3124 | non-null | object |
| 6 | odor | 3124 | non-null | int64 |
| 7 | stalk-shape | 3124 | non-null | int64 |
| 8 | stalk-root | 3124 | non-null | int64 |

* stalk-surface-above-ring 3124 non-null int64

10 stalk-surface-below-ring 3124 non-null int64

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 11 | stalk-color-above-ring | 3124 | non-null | int64 |
| 12 | stalk-color-below-ring | 3124 | non-null | int64 |
| 13 | veil-type | 3124 | non-null | int64 |
| 14 | veil-color | 3124 | non-null | int64 |
| 15 | ring-number | 3124 | non-null | int64 |
| 16 | ring-type | 3124 | non-null | int64 |
| 17 | spore-print-color | 3124 | non-null | int64 |
| 18 | population | 3124 | non-null | int64 |
| 19 | habitat | 3124 | non-null | int64 |
| dtypes: float64(2), int64(16), | | object(2) | |  |
| memory usage: 488.2+ KB | |  |  |  |

* dataset.head()

|  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |  |  |  |  |  |  |  |  |  | **stalk-** |  |
|  | **Unnamed:** | **class** | **cap-** | **cap-** | **cap-** | **bruises** | **odor** | **stalk- stalk- surface-** | | |  |
|  | **0** |  | **shape** | **surface** | **color** |  |  | **shape** | **root** | **above-** |  |
|  |  |  |  |  |  |  |  |  |  | **ring** |  |
|  |  |  |  |  |  |  |  |  |  |  |  |
| **0** | 0 | poisonous | 5.0 | 2 | 4.0 | bruises | 6 | 0 | 3 | 2 |  |
| is there |  |
|  |  |  |  |  |  |  |  |  |  |  |
| **1** | 1 | edible | 5.0 | 2 | 9.0 | bruises | 0 | 0 | 2 | 2 |  |
| is there |  |
|  |  |  |  |  |  |  |  |  |  |  |
| **2** | 2 | edible | 0.0 | 2 | 8.0 | bruises | 3 | 0 | 2 | 2 |  |
| is there |  |
|  |  |  |  |  |  |  |  |  |  |  |
| **3** | 3 | poisonous | 5.0 | 3 | 8.0 | bruises | 6 | 0 | 3 | 2 |  |
| is there |  |
|  |  |  |  |  |  |  |  |  |  |  |
| **4** | 4 | edible | 5.0 | 2 | 3.0 | no | 5 | 1 | 3 | 2 |  |
| bruises |  |
|  |  |  |  |  |  |  |  |  |  |  |

https://colab.research.google.com/drive/1aVPQaYSVGyHYRzai0uhXgqrrE6YPTg4A#scrollTo=LmhvPZuU7wu1&printMode=true 2/4

|  |  |  |
| --- | --- | --- |
| 9/19/21, 10:41 PM | | Lab 05 - CSE422.ipynb - Colaboratory |
|  |  |  |
|  | 1 dataset.isnull().sum() |  |
|  |  |  |
|  | Unnamed: 0 | 0 |
|  | class | 0 |
|  | cap-shape | 0 |
|  | cap-surface | 0 |
|  | cap-color | 0 |
|  | bruises | 0 |
|  | odor | 0 |
|  | stalk-shape | 0 |
|  | stalk-root | 0 |
|  | stalk-surface-above-ring | 0 |
|  | stalk-surface-below-ring | 0 |
|  | stalk-color-above-ring | 0 |
|  | stalk-color-below-ring | 0 |
|  | veil-type | 0 |
|  | veil-color | 0 |
|  | ring-number | 0 |
|  | ring-type | 0 |
|  | spore-print-color | 0 |
|  | population | 0 |
|  | habitat | 0 |
|  | dtype: int64 |  |
|  |  |  |



* from sklearn.preprocessing import LabelEncoder

2 encoder = LabelEncoder()

* dataset['class'] = encoder.fit\_transform(dataset['class'])
* dataset['bruises'] = encoder.fit\_transform(dataset['bruises'])
* dataset.info()

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| <class 'pandas.core.frame.DataFrame'> | | | |  |
| RangeIndex: 3124 entries, 0 to 3123 | | |  |  |
| Data | columns (total 20 columns): | |  |  |
| # | Column | Non-Null Count | | Dtype |
| --- | ------ | -------------- | | ----- |
| 0 | Unnamed: 0 | 3124 | non-null | int64 |
| 1 | class | 3124 | non-null | int64 |
| 2 | cap-shape | 3124 | non-null | float64 |
| 3 | cap-surface | 3124 | non-null | int64 |
| 4 | cap-color | 3124 | non-null | float64 |
| 5 | bruises | 3124 | non-null | int64 |
| 6 | odor | 3124 | non-null | int64 |
| 7 | stalk-shape | 3124 | non-null | int64 |
| 8 | stalk-root | 3124 | non-null | int64 |

* stalk-surface-above-ring 3124 non-null int64

10 stalk-surface-below-ring 3124 non-null int64

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 11 | stalk-color-above-ring | 3124 | non-null | int64 |
| 12 | stalk-color-below-ring | 3124 | non-null | int64 |
| 13 | veil-type | 3124 | non-null | int64 |
| 14 | veil-color | 3124 | non-null | int64 |
| 15 | ring-number | 3124 | non-null | int64 |
| 16 | ring-type | 3124 | non-null | int64 |
| 17 | spore-print-color | 3124 | non-null | int64 |
| 18 | population | 3124 | non-null | int64 |

https://colab.research.google.com/drive/1aVPQaYSVGyHYRzai0uhXgqrrE6YPTg4A#scrollTo=LmhvPZuU7wu1&printMode=true 3/4

|  |  |  |
| --- | --- | --- |
| 9/19/21, 10:41 PM | | Lab 05 - CSE422.ipynb - Colaboratory |
|  |  |  |
|  | 19 habitat | 3124 non-null int64 |
|  | dtypes: float64(2), | int64(18) |
|  | memory usage: 488.2 | KB |
|  |  |  |



* 0s completed at 10:41 PM

https://colab.research.google.com/drive/1aVPQaYSVGyHYRzai0uhXgqrrE6YPTg4A#scrollTo=LmhvPZuU7wu1&printMode=true 4/4