

Practice Labs Mercedes-Benz... - JupyterLab Self learning

https://proxy-6.vocareum.com/lab/tree/Mercedes-Benz Greener Manufacturing /Mercedes-Benz Greener Manufac 90%

File Edit View Run Kernel Tabs Settings Help

Mercedes-Benz Greener Manufacturing /

Filter files by name

Name Last Modified

Mercedes-Benz Greener Manufacturing /

test.csv 10 days ago

train.csv 10 days ago

Mercedes-Benz Greener Manufacturing

Python 3

```
[1]: # Importing Necessary Librarys
```

```
[2]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
%matplotlib inline
```

Loading the datasets

```
[3]: #Training data
train= pd.read_csv("train.csv")
```

```
[4]: #Testing data
test= pd.read_csv("test.csv")
```

```
[5]: train.head()
```

```
[5]:
```

	ID	y	X0	X1	X2	X3	X4	X5	X6	X8	...	X375	X376	X377	X378	X379	X380	X382	X383	X384	X385	
0	0	130.81	k	v	a	t	a	d	u	j	o	...	0	0	1	0	0	0	0	0	0	0
1	6	88.53	k	t	a	v	e	d	y	i	o	...	1	0	0	0	0	0	0	0	0	0
2	7	76.26	a	z	w	n	c	d	x	j	x	...	0	0	0	0	0	0	1	0	0	0
3	9	80.62	a	z	t	n	f	d	x	i	e	...	0	0	0	0	0	0	0	0	0	0
4	13	78.02	a	z	v	n	f	d	h	d	n	...	0	0	0	0	0	0	0	0	0	0

5 rows x 378 columns

```
[6]: train_target =train['y']
```

Simple 0 0 0 Python 3 | Idle Mode: Command Ln 2, Col 20 Mercedes-Benz Greener Manufacturing.ipynb

Practice Labs Mercedes-Benz... - JupyterLab Self learning

https://proxy-6.vocareum.com/lab/tree/Mercedes-Benz Greener Manufacturing /Mercedes-Benz Greener Manufac 90%

File Edit View Run Kernel Tabs Settings Help

Mercedes-Benz Greener Manufacturing /

Filter files by name

Name Last Modified

Mercedes-Benz Greener Manufacturing /

test.csv 10 days ago

train.csv 10 days ago

Mercedes-Benz Greener Manufacturing

Python 3

```
[6]: train_target =train['y']
```

```
[7]: train_df=train.drop(['ID','y'],axis=1)
```

```
[8]: #final train data
train_df.head()
```

```
[8]:
```

	X0	X1	X2	X3	X4	X5	X6	X8	X10	X11	...	X375	X376	X377	X378	X379	X380	X382	X383	X384	X385
0	k	v	a	t	a	d	u	j	o	0	...	0	0	1	0	0	0	0	0	0	0
1	k	t	a	v	e	d	y	i	o	0	...	1	0	0	0	0	0	0	0	0	0
2	a	z	w	n	c	d	x	j	x	0	...	0	0	0	0	0	0	1	0	0	0
3	a	z	t	n	f	d	x	i	e	0	...	0	0	0	0	0	0	0	0	0	0
4	a	z	v	n	f	d	h	d	n	0	...	0	0	0	0	0	0	0	0	0	0

5 rows x 376 columns

```
[9]: test_df=test.drop(['ID'],axis=1)
```

```
[10]: #Final Test data
test_df.head()
```

```
[10]:
```

	X0	X1	X2	X3	X4	X5	X6	X8	X10	X11	...	X375	X376	X377	X378	X379	X380	X382	X383	X384	X385
0	a	z	v	n	f	d	t	a	w	0	...	0	0	0	1	0	0	0	0	0	0
1	t	b	a	i	a	d	b	g	y	0	...	0	0	1	0	0	0	0	0	0	0
2	a	z	v	a	s	f	d	a	j	j	...	0	0	0	1	0	0	0	0	0	0
3	a	z	i	n	f	d	z	i	n	0	...	0	0	0	1	0	0	0	0	0	0

5 rows x 375 columns

Saving started

Simple 0 0 0 Python 3 | Idle Mode: Command Ln 2, Col 20 Mercedes-Benz Greener Manufacturing.ipynb

```
Practice Labs Mercedes-Benz - JupyterLab Self learning
https://proxy-6.vocareum.com/lab/tree/Mercedes-Benz Greener Manufacturing /Mercedes-Benz Greener Manufac 90%

File Edit View Run Kernel Tabs Settings Help

Mercedes-Benz Greener MaX
Code Python 3

2 az v as f d a j j 0 0 ... 0 0 0 1 0 0 0 0 0 0 0
3 az l n f d z l n 0 0 ... 0 0 0 1 0 0 0 0 0 0 0
4 w s as c d y i m 0 0 ... 1 0 0 0 0 0 0 0 0 0 0

5 rows x 376 columns

[11]: print(train_df.shape)
print(train_df.columns)
(4209, 376)
Index(['X0', 'X1', 'X2', 'X3', 'X4', 'X5', 'X6', 'X8', 'X10', 'X11',
      ...,
      'X375', 'X376', 'X377', 'X378', 'X379', 'X380', 'X382', 'X383', 'X384',
      'X385'],
      dtype='object', length=376)

[12]: print(test_df.shape)
print(test_df.columns)
(4209, 376)
Index(['X0', 'X1', 'X2', 'X3', 'X4', 'X5', 'X6', 'X8', 'X10', 'X11',
      ...,
      'X375', 'X376', 'X377', 'X378', 'X379', 'X380', 'X382', 'X383', 'X384',
      'X385'],
      dtype='object', length=376)

Check variance is equal to zero

[13]: from sklearn.feature_selection import VarianceThreshold
variance = VarianceThreshold(threshold=0)

[14]: train_df_removed_zeros = variance.fit_transform(train_df.iloc[:,9:])
train_df_removed_zeros

[14]: array([[0, 1, 0, ..., 0, 0, 0],
          [0, 0, 0, ..., 0, 0, 0],
          [0, 0, 0, ..., 0, 0, 0],
          ...,
          [1, 1, 0, ..., 0, 0, 0],
          [0, 0, 1, ..., 0, 0, 0],
          [0, 0, 0, ..., 0, 0, 0]])

[15]: label_data = train_df.iloc[:,0:8]
label_data.head()

X0 X1 X2 X3 X4 X5 X6 X8
0 k v at a d u j o
1 k t av e d y l o
2 az w n c d x j x
3 az t n f d x l e
4 az v n f d h d n

[16]: label_data.nunique()

X0 47
X1 27
X2 44
X3 7
X4 4
X5 29
X6 12
X8 25
dtype: int64
```

```
Practice Labs Mercedes-Benz - JupyterLab Self learning
https://proxy-6.vocareum.com/lab/tree/Mercedes-Benz Greener Manufacturing /Mercedes-Benz Greener Manufac 90%

File Edit View Run Kernel Tabs Settings Help

Mercedes-Benz Greener MaX
Code Python 3

[14]: train_df_removed_zeros = variance.fit_transform(train_df.iloc[:,9:])
train_df_removed_zeros

[14]: array([[0, 1, 0, ..., 0, 0, 0],
          [0, 0, 0, ..., 0, 0, 0],
          [0, 0, 0, ..., 0, 0, 0],
          ...,
          [1, 1, 0, ..., 0, 0, 0],
          [0, 0, 1, ..., 0, 0, 0],
          [0, 0, 0, ..., 0, 0, 0]])

[15]: label_data = train_df.iloc[:,0:8]
label_data.head()

X0 X1 X2 X3 X4 X5 X6 X8
0 k v at a d u j o
1 k t av e d y l o
2 az w n c d x j x
3 az t n f d x l e
4 az v n f d h d n

[16]: label_data.nunique()

X0 47
X1 27
X2 44
X3 7
X4 4
X5 29
X6 12
X8 25
dtype: int64
```

Practice Labs Mercedes-Benz... - JupyterLab Self learning

https://proxy-6.vocareum.com/lab/tree/Mercedes-Benz Greener Manufacturing /Mercedes-Benz Greener Manufac 90%

File Edit View Run Kernel Tabs Settings Help

Mercedes-Benz Greener ManiX

Python 3

```
X0 12
X8 25
dtype: int64

[17]: from sklearn.preprocessing import LabelEncoder
label = LabelEncoder()

[18]: label_df1=label_data.apply(label.fit_transform)

[19]: label_df1.head()

[19]:
```

	X0	X1	X2	X3	X4	X5	X6	X8
0	32	23	17	0	3	24	9	14
1	32	21	19	4	3	28	11	14
2	20	24	34	2	3	27	9	23
3	20	21	34	5	3	27	11	4
4	20	23	34	5	3	12	3	13

```


[20]: label_df1.var()

[20]: X0    188.741938
X1     72.777974
X2    118.808135
X3      3.027295
X4      0.005461
X5     68.076236
X6      8.508730
X8    49.531868
dtype: float64

[21]: without_zeros_train_df = pd.DataFrame(train_df_removed_zeros)
without_zeros_train_df.head()
```

Practice Labs Mercedes-Benz... - JupyterLab Self learning

https://proxy-6.vocareum.com/lab/tree/Mercedes-Benz Greener Manufacturing /Mercedes-Benz Greener Manufac 90%

File Edit View Run Kernel Tabs Settings Help

Mercedes-Benz Greener ManiX

Python 3

```
dtype: float64

[21]: without_zeros_train_df = pd.DataFrame(train_df_removed_zeros)
without_zeros_train_df.head()

[21]:
```

	0	1	2	3	4	5	6	7	8	9	...	345	346	347	348	349	350	351	352	353	354
0	0	1	0	0	0	0	1	0	0	1	...	0	0	1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	1	0	0	0	...	1	0	0	0	0	0	0	0	0	0
2	0	0	0	0	0	1	0	0	0	0	...	0	0	0	0	0	0	0	1	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	...	0	0	0	0	0	0	0	0	0	0

5 rows × 355 columns

```


[22]: train_df_final = pd.concat([label_df1,without_zeros_train_df],axis=1)
train_df_final.head()

[22]:
```

	X0	X1	X2	X3	X4	X5	X6	X8	0	1	...	345	346	347	348	349	350	351	352	353	354
0	32	23	17	0	3	24	9	14	0	1	...	0	0	1	0	0	0	0	0	0	0
1	32	21	19	4	3	28	11	14	0	...	1	0	0	0	0	0	0	0	0	0	0
2	20	24	34	2	3	27	9	23	0	...	0	0	0	0	0	0	0	1	0	0	0
3	20	21	34	5	3	27	11	4	0	...	0	0	0	0	0	0	0	0	0	0	0
4	20	23	34	5	3	12	3	13	0	...	0	0	0	0	0	0	0	0	0	0	0

5 rows × 363 columns

```


[23]: train_df_final.isna().sum().any()
```

Practice Labs Mercedes-Benz... - JupyterLab Self learning

https://proxy-6.vocareum.com/lab/tree/Mercedes-Benz Greener Manufacturing /Mercedes-Benz Greener Manufac 90%

File Edit View Run Kernel Tabs Settings Help

Mercedes-Benz Greener Manufacturing /

Filter files by name

Name Last Modified

Mercedes-Benz Greener Manufacturing /

test.csv 10 days ago

train.csv 10 days ago

Code Python 3

```
[23]: train_df_final.isna().sum().any()
[23]: False
[24]: test_df.head()
[24]:
```

	X0	X1	X2	X3	X4	X5	X6	X8	X10	X11	...	X375	X376	X377	X378	X379	X380	X382	X383	X384	X385
0	az	v	n	f	d	t	a	w	0	0	...	0	0	0	1	0	0	0	0	0	0
1	t	b	ai	a	d	b	g	y	0	0	...	0	0	1	0	0	0	0	0	0	0
2	az	v	as	f	d	a	j	j	0	0	...	0	0	0	1	0	0	0	0	0	0
3	az	l	n	f	d	z	l	n	0	0	...	0	0	0	1	0	0	0	0	0	0
4	w	s	as	c	d	y	i	m	0	0	...	1	0	0	0	0	0	0	0	0	0

5 rows x 376 columns

```
[25]: test_df.var().sort_values().head(10)
[25]:
```

X295	0.000000	
X369	0.000000	
X296	0.000000	
X257	0.000000	
X258	0.000000	
X278	0.000238	
X233	0.000238	
X280	0.000238	
X290	0.000238	
X293	0.000238	

dtype: float64

```
[26]: test_df_removed_zeros = variance.transform(test_df.iloc[:,9:])
test_df_removed_zeros
```

Practice Labs Mercedes-Benz... - JupyterLab Self learning

https://proxy-6.vocareum.com/lab/tree/Mercedes-Benz Greener Manufacturing /Mercedes-Benz Greener Manufac 90%

File Edit View Run Kernel Tabs Settings Help

Mercedes-Benz Greener Manufacturing /

Filter files by name

Name Last Modified

Mercedes-Benz Greener Manufacturing /

test.csv 10 days ago

train.csv 10 days ago

Code Python 3

```
dtype: float64
[26]: test_df_removed_zeros = variance.transform(test_df.iloc[:,9:])
test_df_removed_zeros
[26]:
```

0	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	0	0	0	...	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

5 rows x 355 columns

```
[27]: without_zeros_test_df=pd.DataFrame(test_df_removed_zeros)
without_zeros_test_df.head()
[27]:
```

	0	1	2	3	4	5	6	7	8	9	...	345	346	347	348	349	350	351	352	353	354	
0	0	0	0	0	0	0	0	0	0	0	...	0	0	0	1	0	0	0	0	0	0	0
1	0	0	0	0	0	0	0	1	0	0	...	0	0	1	0	0	0	0	0	0	0	0
2	0	0	1	0	0	0	0	0	0	0	...	0	0	0	1	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0	0	0	0	...	0	0	0	1	0	0	0	0	0	0	0
4	0	0	1	0	0	0	0	0	0	0	...	1	0	0	0	0	0	0	0	0	0	0

5 rows x 355 columns

```
[28]: label_data2=test_df.iloc[:,0:8]
label_data.head()
[28]:
```

	X0	X1	X2	X3	X4	X5	X6	X8
0	k	v	a	t	a	d	u	j
1	k	t	a	v	e	d	y	l

Practice Labs Mercedes-Ben... - JupyterLab Self learning

https://proxy-6.vocareum.com/lab/tree/Mercedes-Benz Greener Manufacturing /Mercedes-Benz Greener Manufac 90%

File Edit View Run Kernel Tabs Settings Help

Mercedes-Benz Greener Manufacturing /

Filter files by name

Name Last Modified

Mercedes-... a minute ago

test.csv 10 days ago

train.csv 10 days ago

Mercedes-Benz Greener Manufacturing

5 rows x 355 columns

```
[28]: label_data2=test_df.iloc[:,0:8]
      label_data.head()
```

```
[28]:
```

	X0	X1	X2	X3	X4	X5	X6	X8
0	k	v	a	t	a	d	u	j
1	k	t	a	v	e	d	y	i
2	a	z	w	n	c	d	x	j
3	a	z	t	n	f	d	x	i
4	a	z	v	n	f	d	h	d

```
[29]: label_df3=label_data2.apply(label_fit_transform)
```

```
[30]: label_df3.head()
```

```
[30]:
```

	X0	X1	X2	X3	X4	X5	X6	X8
0	21	23	34	5	3	26	0	22
1	42	3	8	0	3	9	6	24
2	21	23	17	5	3	0	9	9
3	21	13	34	5	3	31	11	13
4	45	20	17	2	3	30	8	12

```
[31]: test_df_final = pd.concat([label_df3,without_zeros_test_df],axis=1)
      test_df_final.head()
```

```
[31]:
```

	X0	X1	X2	X3	X4	X5	X6	X8	0	1	...	345	346	347	348	349	350	351	352	353	354
0	21	23	34	5	3	26	0	22	0	0	...	345	346	347	348	349	350	351	352	353	354
1	42	3	8	0	3	9	6	24	0	0	...	345	346	347	348	349	350	351	352	353	354
2	21	23	17	5	3	0	9	9	0	0	...	345	346	347	348	349	350	351	352	353	354
3	21	13	34	5	3	31	11	13	0	0	...	345	346	347	348	349	350	351	352	353	354
4	45	20	17	2	3	30	8	12	0	0	...	345	346	347	348	349	350	351	352	353	354

Simple 0 0 Python 3 | Idle Mode: Command Ln 2, Col 20 Mercedes-Benz Greener Manufacturing.ipynb

Practice Labs Mercedes-Ben... - JupyterLab Self learning

https://proxy-6.vocareum.com/lab/tree/Mercedes-Benz Greener Manufacturing /Mercedes-Benz Greener Manufac 90%

File Edit View Run Kernel Tabs Settings Help

Mercedes-Benz Greener Manufacturing /

Filter files by name

Name Last Modified

Mercedes-... a minute ago

test.csv 10 days ago

train.csv 10 days ago

Mercedes-Benz Greener Manufacturing

```
[31]: test_df_final = pd.concat([label_df3,without_zeros_test_df],axis=1)
      test_df_final.head()
```

```
[31]:
```

	X0	X1	X2	X3	X4	X5	X6	X8	0	1	...	345	346	347	348	349	350	351	352	353	354
0	21	23	34	5	3	26	0	22	0	0	...	345	346	347	348	349	350	351	352	353	354
1	42	3	8	0	3	9	6	24	0	0	...	345	346	347	348	349	350	351	352	353	354
2	21	23	17	5	3	0	9	9	0	0	...	345	346	347	348	349	350	351	352	353	354
3	21	13	34	5	3	31	11	13	0	0	...	345	346	347	348	349	350	351	352	353	354
4	45	20	17	2	3	30	8	12	0	0	...	345	346	347	348	349	350	351	352	353	354

5 rows x 363 columns

Perform dimensionality reduction.

```
[32]: from sklearn.decomposition import PCA
      pca = PCA(n_components = 0.2)
```

```
[33]: # Train and Test split on Train dataset
      from sklearn.model_selection import train_test_split
```

```
[34]: x_train,x_test,y_train,y_test = train_test_split(train_df_final ,train_target,test_size=0.3,random_state=42)
```

```
[35]: x_train.shape,x_test.shape,y_train.shape,y_test.shape
```

```
[35]: ((2946, 363), (1263, 363), (2946,), (1263,))
```

```
[36]: x_train =pca.fit_transform(x_train)
      x_test = pca.transform(x_test)
```

/usr/local/lib/python3.7/site-packages/sklearn/utils/validation.py:1692: FutureWarning: Feature names only support names that are all strings. Got feature names with dtypes: ['int', 'str']. An error will be raised in 1.2.

Simple 0 0 Python 3 | Idle Mode: Command Ln 2, Col 20 Mercedes-Benz Greener Manufacturing.ipynb

Practice Labs Mercedes-Benz... - JupyterLab Self learning

https://proxy-6.vocareum.com/lab/tree/Mercedes-Benz Greener Manufacturing /Mercedes-Benz Greener Manufac 90%

File Edit View Run Kernel Tabs Settings Help

Mercedes-Benz Greener Manufacturing /

Filter files by name

Name Last Modified

Mercedes-Benz Greener Manufacturing /

test.csv 10 days ago

train.csv 10 days ago

Code Python 3

```
x_test = pca.transform(x_test)

/usr/local/lib/python3.7/site-packages/sklearn/utils/validation.py:1692: FutureWarning: Feature names only support names that are all strings. Got feature names with dtypes: ['int', 'str']. An error will be raised in 1.2.
FutureWarning,
/usr/local/lib/python3.7/site-packages/sklearn/utils/validation.py:1692: FutureWarning: Feature names only support names that are all strings. Got feature names with dtypes: ['int', 'str']. An error will be raised in 1.2.
FutureWarning,

[37]: test_df=pca.transform(test_df_final)

/usr/local/lib/python3.7/site-packages/sklearn/utils/validation.py:1692: FutureWarning: Feature names only support names that are all strings. Got feature names with dtypes: ['int', 'str']. An error will be raised in 1.2.
FutureWarning,

[38]: pca.n_components_

[38]: 1

[39]: pca.explained_variance_ratio_

[39]: array([0.38296186])

XGBoost

[40]: from sklearn import svm
from sklearn.metrics import r2_score,mean_squared_error
from xgboost import XGBRegressor

[41]: xgbreg = XGBRegressor(random_state=42)

[42]: model = xgbreg.fit(x_train,y_train)

[43]: ypred_train = model.predict(x_train)
ypred_train

[43]: array([ 97.8411 , 105.01268 , 99.95553 , ..., 93.519264 , 94.643814,
98.87075 ], dtype=float32)

[44]: ypred_test = model.predict(x_test)
ypred_test

[44]: array([ 94.144424, 104.580864, 104.58693 , ..., 98.87075 , 101.180534,
102.3027 ], dtype=float32)

[45]: print(r2_score(ypred_train,y_train))

0.06373876967202363

[46]: print(mean_squared_error(ypred_train,y_train))

54.37626614461824

[47]: prediction = pd.DataFrame({'ytest ':y_test,'ypred':ypred_test})

[48]: prediction

[48]:
```

	ytest	ypred
1073	97.94	94.144424
144	96.41	104.580864
2380	105.83	104.586929
184	79.09	80.366035
2587	108.69	111.047722
...

Simple 0 0 Python 3 | Idle Mode: Command Ln 2, Col 20 Mercedes-Benz Greener Manufacturing.ipynb

Practice Labs Mercedes-Benz... - JupyterLab Self learning

https://proxy-6.vocareum.com/lab/tree/Mercedes-Benz Greener Manufacturing /Mercedes-Benz Greener Manufac 90%

File Edit View Run Kernel Tabs Settings Help

Mercedes-Benz Greener Manufacturing /

Filter files by name

Name Last Modified

Mercedes-Benz Greener Manufacturing /

test.csv 10 days ago

train.csv 10 days ago

Code Python 3

```
[42]: model = xgbreg.fit(x_train,y_train)

[43]: ypred_train = model.predict(x_train)
ypred_train

[43]: array([ 97.8411 , 105.01268 , 99.95553 , ..., 93.519264 , 94.643814,
98.87075 ], dtype=float32)

[44]: ypred_test = model.predict(x_test)
ypred_test

[44]: array([ 94.144424, 104.580864, 104.58693 , ..., 98.87075 , 101.180534,
102.3027 ], dtype=float32)

[45]: print(r2_score(ypred_train,y_train))

0.06373876967202363

[46]: print(mean_squared_error(ypred_train,y_train))

54.37626614461824

[47]: prediction = pd.DataFrame({'ytest ':y_test,'ypred':ypred_test})

[48]: prediction

[48]:
```

	ytest	ypred
1073	97.94	94.144424
144	96.41	104.580864
2380	105.83	104.586929
184	79.09	80.366035
2587	108.69	111.047722
...

Simple 0 0 Python 3 | Idle Mode: Command Ln 2, Col 20 Mercedes-Benz Greener Manufacturing.ipynb

Practice LabsMercedes-Ben... - JupyterLabSelf learning

https://proxy-6.vocareum.com/lab/tree/Mercedes-Benz Greener Manufacturing /Mercedes-Benz Greener Manufac90%

File Edit View Run Kernel Tabs Settings Help

Filter files by name

Mercedes-Benz Greener Manufacturing /

Mercedes-...2 minutes ago

test.csv10 days ago

train.csv10 days ago

Mercedes-Benz Greener MaX

54.37626614461824

[47]:

prediction = pd.DataFrame({'ytest':y_test,'ypred':ypred_test})

[48]:

prediction

[48]:

	ytest	ypred
1073	97.94	94.144424
144	96.41	104.580864
2380	105.83	104.586929
184	79.09	80.366035
2587	108.69	111.047722
...
2493	115.25	98.354218
3388	88.59	101.401352
3997	92.90	98.870750
383	98.24	101.180534
3364	91.46	102.302696

1263 rows x 2 columns

[]:

Simple0Python 3 | Idle

Mode: CommandLn 2, Col 20Mercedes-Benz Greener Manufacturing.ipynb