

CAR RESALE VALUE PREDICTION

import libraries

```
import pandas as pd
import numpy as np
import matplotlib as plt
from sklearn.preprocessing import LabelEncoder
import pickle
```

read the dataset

```
df = pd.read_csv("Data/autos.csv", header=0, sep=',', encoding='Latin1')
df
```

preprocess dataset

```
df=df.drop('offerType',axis=1)
df=df[(df.powerPS > 50) & (df.powerPS < 900)]
df = df[(df.yearOfRegistration >= 1950) & (df.yearOfRegistration < 2017)]
df.drop(['name', 'abtest', 'dateCrawled', 'nrOfPictures', 'lastSeen', 'postalCode','dateCreated'],
axis='columns',inplace=True)
new_df = df.copy()
```

```
new_df = new_df.drop_duplicates ([ 'price', 'vehicleType', 'yearOfRegistration', 'gearbox',
'powerPS', 'model', 'kilometer', 'monthOfRegistration', 'fuelType', 'notRepairedDamage'])
```

```
new_df.gearbox.replace(('manuell', 'automatik'), ('manual', 'automatic'), inplace=True)
```

```
new_df.fuelType.replace(('benzin', 'andere', 'elektro'), ('petrol', 'others', 'electric'), inplace=True)
```

```
new_df.vehicleType.replace(('kleinwagen', 'cabrio', 'komb', 'andere'), ('small car', 'convertible',
'combination', 'others'), inplace=True) new_df.notRepairedDamage.replace(('ja', 'nein'), ('Yes',
'No'),inplace=True)
```

```
new_df = new_df[(new_df.price >= 100) & (new_df.price <= 150000)]
```

```
new_df['notRepairedDamage'].fillna(value='not-declared', inplace=True)
```

```
new_df['fuelType'].fillna(value='not-declared', inplace=True) new_df['gearbox'].fillna(value='not-
```

```
declared', inplace=True) new_df['vehicleType'].fillna (value='not-declared', inplace=True)
new_df['model'].fillna(value='not-declared',inplace=True)

new_df.to_csv("autos_preprocessed.csv")
```