Feature Selection Techniques in Machine Learning with Python

Benefits of performing feature selection:

- **Improves Accuracy:** Less misleading data means modeling accuracy improves.
- **Reduces Training Time:** fewer data points reduce algorithm complexity and algorithms train faster.
- **Reduces Overfitting:** Less redundant data means less opportunity to make decisions based on noise

Feature Selection Methods

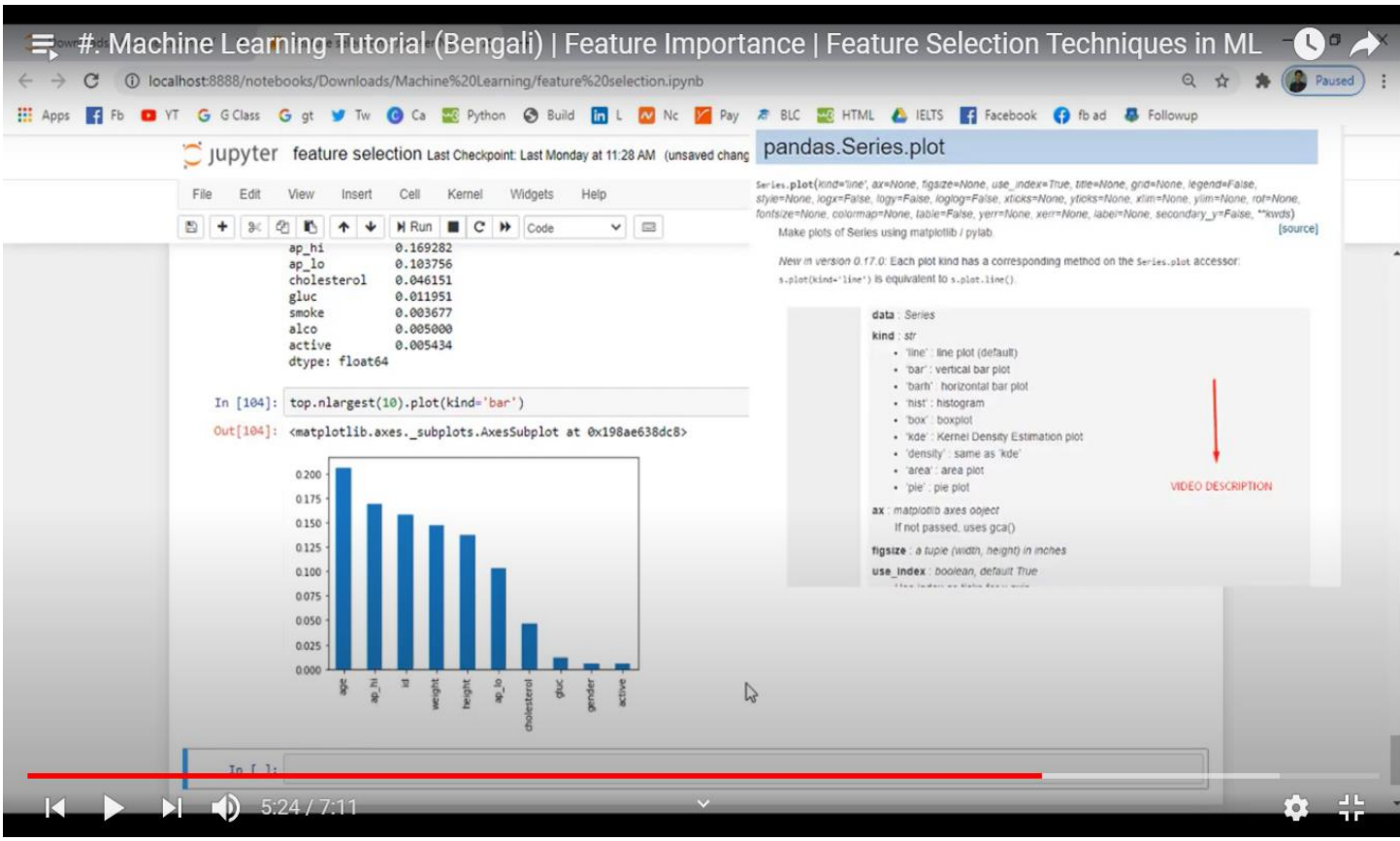
1. Univariate Selection
2. Feature Importance
3. Correlation Matrix with Heat map

Univariate Selection

Score Functions:

- chi-square
- f_classif
- mutual_info_classif

Note: all links available in video description


#. Machine Learning Tutorial (Bengali) | Feature Importance | Feature Selection Techniques in ML
localhost:8888/notebooks/Downloads/Machine%20Learning/feature%20selection.ipynb
jupyter feature selection Last Checkpoint: Last Monday at 11:28 AM (unsaved chang
File Edit View Insert Cell Kernel Widgets Help
ap_hi 0.169282
ap_lo 0.103756
cholesterol 0.046151
gluc 0.011951
smoke 0.003677
alco 0.005000
active 0.005434
dtype: float64
In [104]: top.nlargest(10).plot(kind='bar')
Out[104]: <matplotlib.axes._subplots.AxesSubplot at 0x198ae638dc8>
0.200
0.175
0.150
0.125
0.100
0.075
0.050
0.025
0.000
age
ap_hi
id
weight
height
ap_lo
cholesterol
gluc
gender
active
pandas.Series.plot
Make plots of Series using matplotlib / pylab.
[source]
data : Series
kind : str
'line' : line plot (default)
'bar' : vertical bar plot
'barh' : horizontal bar plot
'hist' : histogram
'box' : boxplot
'kde' : Kernel Density Estimation plot
'density' : same as 'kde'
'area' : area plot
'pie' : pie plot
ax : matplotlib axes object
If not passed, uses gca()
figsize : a tuple (width, height) in inches
use_index : boolean, default True
VIDEO DESCRIPTION
5:24 / 7:11