



Hochschule
Bielefeld
University of
Applied Sciences
and Arts

Pratikum Blatt 4 (5 Punkte)

mit dem iris-Datensatz (iris.csv) und mit der Datei „pvttest.csv“

Datenanalyse und Einführung in Maschinelles Lernen
WS 2025/26

Datenanalyse mit Python
Abgabe am 26./27.November 2025

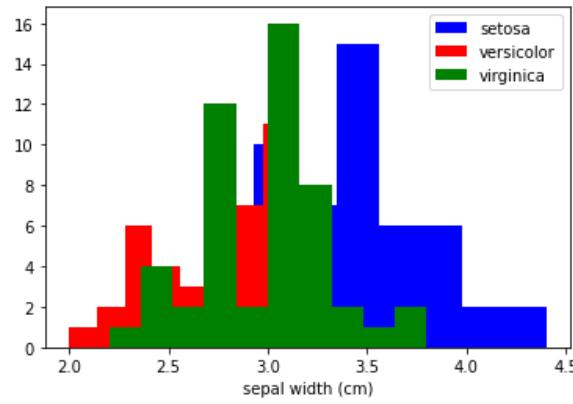
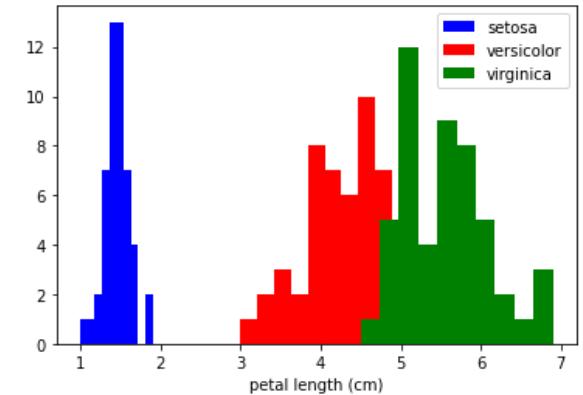
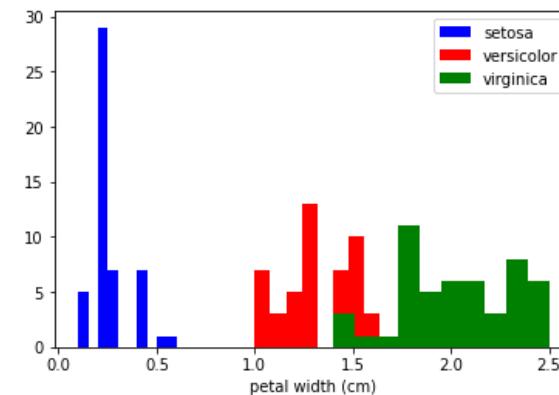
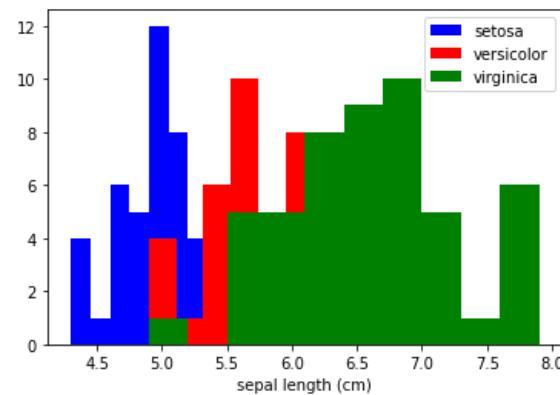
Dozentin: Grit Behrens
mailto: grit.behrens@hsbi.de

Studiengang Informatik Fachbereich Campus Minden



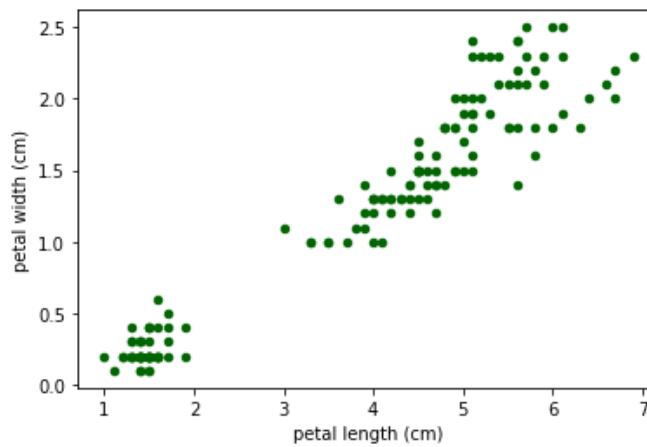
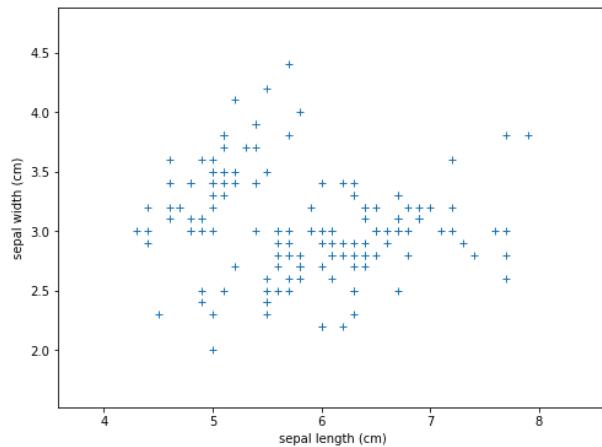
Aufgabe 1 (1 Punkt)

Laden Sie den iris-Datensatz und zeichnen Sie vier Histogramme jeweils der vier numerischen Merkmale für die drei Objektklassen setosa, versicolor, und virginica. Färben Sie nach den Objektklassen ein. Erzeugen Sie die folgenden Ausgaben:



Aufgabe 2 (1 Punkte)

- a) Zeichnen Sie ein Streudiagramm für die Attribute Kelchblatlänge in der x-Achse und Kelchblattbreite in der y-Achse, wie unten in der Abbildung sowie für Blütenblatlänge und Blütenblattbreite auch wie unten in der Abbildung. (0.5 Punkt)



- b) Berechnen Sie den Korrelationskoeffizient für die beiden Merkmalspaare aus Aufgabe 1 (Kelchblatlänge u. Kelchblattbreite, Blütenblatlänge und Blütenblattbreite. Erzeugen Sie die Ausgaben (0.5 Punkt) :

The correlation coefficient between petal length (cm) and petal width (cm) attributes is 0.96287.

The correlation coefficient between petal length (cm) and petal width (cm) attributes is -0.11757.

Aufgabe 3 (3 Punkte)

Lesen Sie den Datensatz „pvtest.csv“ als DataFrame ein. Der Datensatz stammt aus dem Monitoring von einer Photovoltaikanlage und enthält u.a. Werte zur erzielten Stromstärke Dci, zur Spannung Dcu, zur Stringleistung Dcp und zur Temperatur des Wechselrichters in einem Strang von 21 Photovoltaikmodulen der Firma SunTech Power.

- a. Zeigen Sie die ersten 10 Zeilen des DataFrame an (Screenshot beinhaltet nicht alle Spalten) 1P!

Unnamed: 0		Time	Inverter	AcFreq	AcPower	Edaily	Eint	Etotal	Riso	Error	...	City_plant	Street_plant	Latitude_plant	Longitude_plant	ModulArea
0	232	20.01.21 07:45	WR 3 Diehl	NaN	0.0	0.0	NaN	NaN	NaN	NaN	...	Rüsselsheim	Lenbachstr. 17 - 27	49.988792	8.442296	4
1	233	20.01.21 07:50	WR 3 Diehl	NaN	0.0	0.0	NaN	NaN	NaN	NaN	...	Rüsselsheim	Lenbachstr. 17 - 27	49.988792	8.442296	4
2	234	20.01.21 07:55	WR 3 Diehl	NaN	0.0	0.0	NaN	NaN	NaN	NaN	...	Rüsselsheim	Lenbachstr. 17 - 27	49.988792	8.442296	4
3	235	20.01.21 08:00	WR 3 Diehl	NaN	0.0	0.0	NaN	NaN	NaN	NaN	...	Rüsselsheim	Lenbachstr. 17 - 27	49.988792	8.442296	4
4	236	20.01.21 08:05	WR 3 Diehl	NaN	0.0	0.0	NaN	NaN	NaN	NaN	...	Rüsselsheim	Lenbachstr. 17 - 27	49.988792	8.442296	4
5	237	20.01.21 08:15	WR 3 Diehl	NaN	0.0	0.0	NaN	NaN	NaN	NaN	...	Rüsselsheim	Lenbachstr. 17 - 27	49.988792	8.442296	4
6	238	20.01.21 08:20	WR 3 Diehl	NaN	0.0	0.0	NaN	NaN	NaN	NaN	...	Rüsselsheim	Lenbachstr. 17 - 27	49.988792	8.442296	4
7	239	20.01.21 08:25	WR 3 Diehl	NaN	0.0	0.0	NaN	NaN	NaN	NaN	...	Rüsselsheim	Lenbachstr. 17 - 27	49.988792	8.442296	4
8	240	20.01.21 08:30	WR 3 Diehl	NaN	0.0	0.0	NaN	NaN	NaN	NaN	...	Rüsselsheim	Lenbachstr. 17 - 27	49.988792	8.442296	4
9	241	20.01.21 08:35	WR 3 Diehl	NaN	11.0	0.0	NaN	NaN	NaN	NaN	...	Rüsselsheim	Lenbachstr. 17 - 27	49.988792	8.442296	4

10 rows x 54 columns

Aufgabe 3

b) Lassen Sie sich die Namen der Spalten anzeigen, selektieren Sie dann die Spalten mit den Namen ‚Dci‘, ‚Dcp‘, ‚Dcu‘ und ‚Temp1‘! Zeigen Sie die ersten 10 Zeilen an (0.5P):

	Dci	Dcp	Dcu	Temp1
0	0.0	0	0	0.0
1	0.0	0	0	0.0
2	0.0	0	0	0.0
3	0.0	0	0	0.0
4	0.0	0	0	0.0
5	0.0	0	0	0.0
6	0.0	0	0	0.0
7	0.0	0	0	0.0
8	0.0	0	158	3.0
9	0.0	0	251	6.0

c. Filtern Sie alle Zeilen im DataFrame heraus, in denen die Werte ‚Dci‘, Dcp, Dc oder Temp1 gleich Null ist. Überprüfen Sie das Ergebnis und zeigen Sie die 10 ersten Zeilen an (0.5P):

	Dci	Dcp	Dcu	Temp1
10	0.005391	2	371	6.0
11	0.021505	10	465	6.0
12	0.034553	17	492	7.0
13	0.062147	33	531	8.0
14	0.088929	49	551	8.0
15	0.106572	60	563	8.0
16	0.117958	67	568	8.0
17	0.133100	76	571	8.0
18	0.157534	92	584	9.0
19	0.177172	104	587	10.0

Aufgabe 3

d. Erstellen Sie eine Heatmap über die verbliebenen 4 Merkmale wie unten abgebildet! (1P)

