REC	Age	Income	Student	Credit_rating	Buys_computer
r1	<=30	High	No	Fair	No
r2	<=30	High	No	Excellent	No
r3	31..40	High	No	Fair	Yes
r4	>40	Medium	No	Fair	Yes
r5	>40	Low	Yes	Fair	Yes
r6	>40	Low	Yes	Excellent	No
r7	31...40	Low	Yes	Excellent	Yes
r8	<=30	Medium	No	Fair	No
r9	<=30	Low	Yes	Fair	Yes
r10	>40	Medium	Yes	Fair	Yes
r11	<=30	Medium	Yes	Excellent	Yes
r12	31...40	Medium	No	Excellent	Yes
r13	31...40	High	Yes	Fair	Yes
r14	>40	Medium	No	Excellent	No
r15	<=30	Medium	No	Excellent	No
r16	<=30	Low	No	Fair	No
r17	<=30	Low	No	Excellent	No
r18	31...40	Low	Yes	Fair	Yes
r19	>40	Medium	Yes	Excellent	Yes
r20	31...40	High	No	Excellent	Yes

Aufgabe A

Gesucht: Entscheidung ob jemand einen Computer kauft ja oder nein.

$$E(S)=\sum_{i=1}^{c} -p(i)*(\log_2 * p(i))$$

Kauft Computer	
ja	nein
12	8

Entropie = E

E(KaufComputer) = E(12,8)

= E(0.6,0.4)

= - (0.6 $\log_2(0.6)$) - (0.4 $\log_2(0.4)$)

= 0.97095059

0.001609492

0.224371171

0.03030514

Zugewinn(T,X) = Z(T,X) = Entropie(T) – Entropie(T,X)

		Kauft Computer ja	nein	
Alter	<=30	2	6	0.32451125
	31-40	6	0	
	>40	4	2	+ 0.27548875
Z(KaufComputer, Alter) =				= 0.6
E(KaufComputer) – E(KaufComputer,Alter) =				
0.97095059 – 0.6 = 0.37095059				

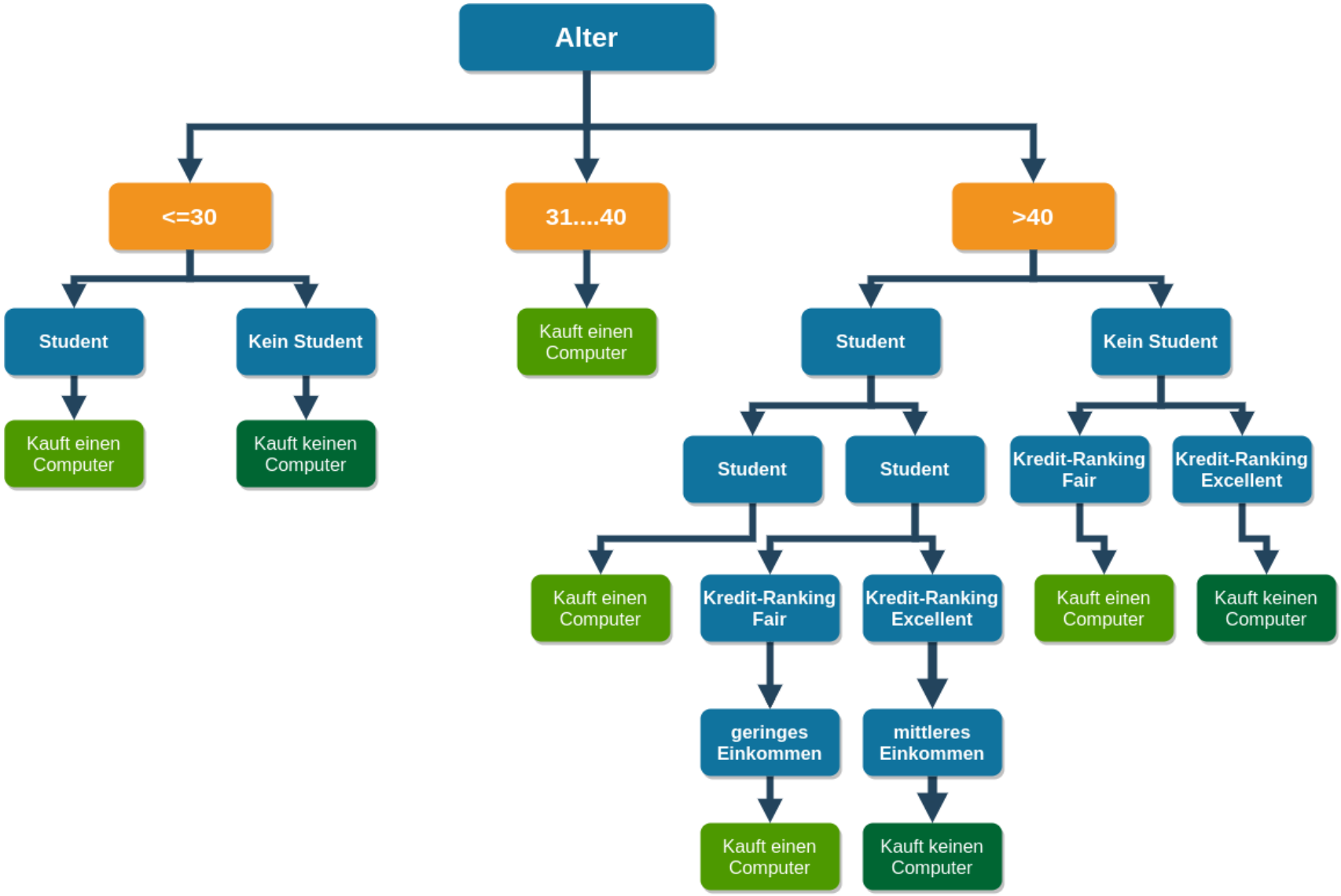
		Kauft Computer ja	nein	
Einkommen	Niedrig	4	3	0.344829848
	Mittel	5	3	+ 0.381773601
	Hoch	3	2	+ 0.242737649
Z(KaufComputer, Einkommen) =				= 0.969341098
E(KaufComputer) – E(KaufComputer,Einkommen) =				
0.97095059 – 0.969341098 = 0.00169492				

		Kauft Computer ja	nein	
Student	ja	8	1	0.226466251
	nein	4	7	+ 0.520113168
Z(KaufComputer, Student) =				= 0.746579419
E(KaufComputer) – E(KaufComputer,Student) =				
0.97095059 – 0.746579419 = 0.224371171				

		Kauft Computer ja	nein	
Kredit-Rating	Fair	7	3	0.44064545
	Excellent	5	5	+ 0.5
Z(KaufComputer, Kredit-Rating) =				= 0.94064545
E(KaufComputer) – E(KaufComputer,Kredit-Rating) =				
0.97095059 – 0.94064545 = 0.03030514				

Aus den Berechnungen ergibt sich die Reihenfolge: Alter, Student, Kredit-Ranking, Einkommen

Daraus ergibt sich dann nachfolgender Entscheidungsbau:


Alter
<=30
31....40
>40
Student
Kein Student
Kauft einen Computer
Kauft keinen Computer
Kauft einen Computer
Student
Kein Student
Student
Student
Kredit-Ranking Fair
Kredit-Ranking Excellent
Kauft einen Computer
Kredit-Ranking Fair
Kredit-Ranking Excellent
Kauft einen Computer
Kauft keinen Computer
geringes Einkommen
mittleres Einkommen
Kauft einen Computer
Kauft keinen Computer

Aufgabe B

In [1]:
```
import pandas as pd
import numpy as np

from sklearn.tree import DecisionTreeClassifier
from sklearn import preprocessing

data = pd.read_csv("dataset.csv")
data = data.drop('REC', axis=1)
data2 = pd.read_csv("dataset.csv")
data2 = data2.drop('REC', axis=1)
le = preprocessing.LabelEncoder()
le.fit(["<=30","31...40",">40","Low","Medium","High","Fair","Excellent","Yes","No"])
data = data.values.reshape(-1,1)
new_values = le.transform(data).reshape(-1,5)
```

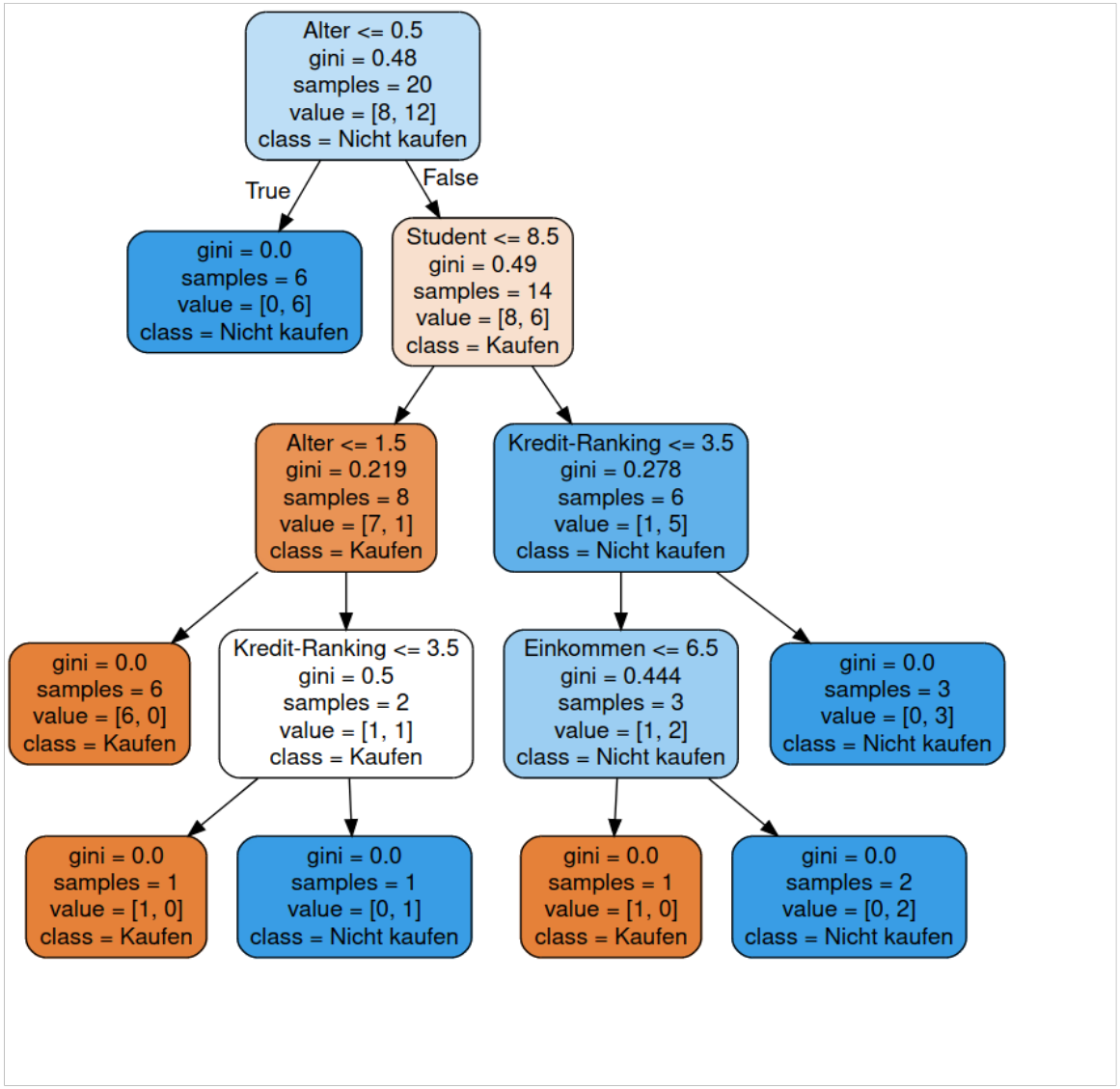
In [2]:
```
X = new_values[:, 0:4] # petal length and width
y = new_values[:, 4:5]

tree_clf = DecisionTreeClassifier(max_depth=8)
tree_clf.fit(X, y)
```

Out[2]:
```
DecisionTreeClassifier(max_depth=8)
```

In [3]:
```
from sklearn.tree import export_graphviz

export_graphviz(
        tree_clf,
    out_file="/media/max/PROJECTS/coden/Data Science Modul/MM_11_Decision_Trees_Exercise/test.dot",
        feature_names=data2.columns.values[0:4].tolist(),
        class_names=["Kaufen", "Nicht kaufen"],
        rounded=True,
        filled=True
)
```


Alter <= 0.5
gini = 0.48
samples = 20
value = [8, 12]
class = Nicht kaufen
True
False
gini = 0.0
samples = 6
value = [0, 6]
class = Nicht kaufen
Student <= 8.5
gini = 0.49
samples = 14
value = [8, 6]
class = Kaufen
Alter <= 1.5
gini = 0.219
samples = 8
value = [7, 1]
class = Kaufen
Kredit-Ranking <= 3.5
gini = 0.278
samples = 6
value = [1, 5]
class = Nicht kaufen
gini = 0.0
samples = 6
value = [6, 0]
class = Kaufen
Kredit-Ranking <= 3.5
gini = 0.5
samples = 2
value = [1, 1]
class = Kaufen
Einkommen <= 6.5
gini = 0.444
samples = 3
value = [1, 2]
class = Nicht kaufen
gini = 0.0
samples = 3
value = [0, 3]
class = Nicht kaufen
gini = 0.0
samples = 1
value = [1, 0]
class = Kaufen
gini = 0.0
samples = 1
value = [0, 1]
class = Nicht kaufen
gini = 0.0
samples = 1
value = [1, 0]
class = Kaufen
gini = 0.0
samples = 2
value = [0, 2]
class = Nicht kaufen