

## Aufgabe 1:

Unter Verwendung des bereitgestellten JAVA-Projekts wurden die eigenen Messdaten (Thymio wurde während der Messungen etwas bewegt, um größere Streuung der Werte zu erreichen – realistischere Werte) in eine .arff Datei geschrieben. Klassifikation in Weka:

The image displays two screenshots of the Weka Explorer application, showing the results of a classification task using different models.

**Top Screenshot: NaiveBayes Classifier**

**Classifier:** NaiveBayes

**Test options:** Cross-validation (Folds: 10)

**Classifier output:**

Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	2235	99.3333 %
Incorrectly Classified Instances	15	0.6667 %
Kappa statistic	0.99	
Mean absolute error	0.0414	
Root mean squared error	0.1311	
Relative absolute error	9.3239 %	
Root relative squared error	27.8133 %	
Total Number of Instances	2250	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.999	0	1	0.999	0.999	1	ECKE
	0.981	0	1	0.981	0.991	1	FRONTAL
	1	0.01	0.98	1	0.99	0.996	KANTE
Weighted Avg.	0.993	0.003	0.993	0.993	0.993	0.999	

=== Confusion Matrix ===

a	b	c	<-- classified as
749	0	1	a = ECKE
0	736	14	b = FRONTAL
0	0	750	c = KANTE

**Bottom Screenshot: RandomForest Classifier**

**Classifier:** RandomForest -I 100-K 0-S 1

**Test options:** Cross-validation (Folds: 10)

**Classifier output:**

Time taken to build model: 0.43 seconds

=== Stratified cross-validation ===

=== Summary ===

Correctly Classified Instances	2250	100 %
Incorrectly Classified Instances	0	0 %
Kappa statistic	1	
Mean absolute error	0.0001	
Root mean squared error	0.0071	
Relative absolute error	0.0327 %	
Root relative squared error	1.5051 %	
Total Number of Instances	2250	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	1	0	1	1	1	1	ECKE
	1	0	1	1	1	1	FRONTAL
	1	0	1	1	1	1	KANTE
Weighted Avg.	1	0	1	1	1	1	

=== Confusion Matrix ===

a	b	c	<-- classified as
750	0	0	a = ECKE
0	750	0	b = FRONTAL
0	0	750	c = KANTE

Das Verhältnis aus Anzahl der als Klasse A klassifizierten Daten und Anzahl der korrekt als A klassifizierten Daten wird hier als Precision bezeichnet.

## Aufgabe 2:

Zur Diskretisierung der Daten wurde die entsprechende Methode im bereitgestellten JAVA-Projekt implementiert. Für die Längendiskretisierung wurde der längste Abstand ermittelt; der kürzere Abstand wurde dann auf diesen normalisiert. Die Quartile der Winkel wurden in R über den summary-Befehl ermittelt, jeweils für linke und rechte Winkel.

## Aufgabe 3:

Ergebnisse der Klassifizierung der diskretisierten Daten (anscheinend schlechtere Ergebnisse als Rohdaten):

The image displays two screenshots of the Weka Explorer application, showing the results of classification tasks using different models.

**Top Screenshot: NaiveBayes Classifier**

**Classifier:** NaiveBayes

**Test options:** Cross-validation (Folds: 10)

**Classifier output:**

Time taken to build model: 0 seconds

=== Stratified cross-validation ===

=== Summary ===

Metric	Value	Percentage
Correctly Classified Instances	2209	98.1778 %
Incorrectly Classified Instances	41	1.8222 %
Kappa statistic	0.9727	
Mean absolute error	0.0134	
Root mean squared error	0.1077	
Relative absolute error	3.021 %	
Root relative squared error	22.8364 %	
Total Number of Instances	2250	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	0.976	0.003	0.993	0.976	0.985	0.985	ECKE
	0.981	0.011	0.977	0.981	0.979	0.986	FRONTAL
	0.988	0.013	0.975	0.988	0.981	0.991	KANTE
Weighted Avg.	0.982	0.009	0.982	0.982	0.982	0.987	

=== Confusion Matrix ===

	a	b	c	<-- classified as
732	13	5		a = ECKE
0	736	14		b = FRONTAL
5	4	741		c = KANTE

**Bottom Screenshot: RandomForest Classifier**

**Classifier:** RandomForest -1 100-K 0-5 1

**Test options:** Cross-validation (Folds: 10)

**Classifier output:**

Time taken to build model: 0.36 seconds

=== Stratified cross-validation ===

=== Summary ===

Metric	Value	Percentage
Correctly Classified Instances	2245	99.7778 %
Incorrectly Classified Instances	5	0.2222 %
Kappa statistic	0.9967	
Mean absolute error	0.0023	
Root mean squared error	0.0362	
Relative absolute error	0.512 %	
Root relative squared error	7.6702 %	
Total Number of Instances	2250	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	ROC Area	Class
	1	0.002	0.996	1	0.998	1	ECKE
	1	0.001	0.997	1	0.999	1	FRONTAL
	0.993	0	1	0.993	0.997	0.999	KANTE
Weighted Avg.	0.998	0.001	0.998	0.998	0.998	1	

=== Confusion Matrix ===

	a	b	c	<-- classified as
750	0	0		a = ECKE
0	750	0		b = FRONTAL
3	2	745		c = KANTE

#### **Aufgabe 4:**

Aufgrund der Ergebnisse in Weka wurde RandomForest als Klassifikator gewählt, als Datensatz zunächst Messdaten aus idealen Bedingungen, d.h. keine Bewegung des Thymio und ideale Positionen für Frontal, Ecke und Kante. Diese Messdaten wurden in Weka sowohl mit NaiveBayes als auch mit RandomForest 100% richtig klassifiziert. Beim „Livebetrieb“ wurden jedoch alle Positionen außer eindeutig links, rechts oder frei mit 90% Wahrscheinlichkeit als Spitze, 10% als Ecke und 0% als Frontal klassifiziert, selbst für „ideale“ Positionen. Deshalb wurden neue Messdaten gesammelt, diesmal mit realistischeren Bedingungen: Während der Messung wurde der Thymio etwas bewegt, um eine größere Streuung der Werte zu erreichen.

#### **Aufgabe 5:**

(Mit „allen Messungen“ sind Messungen gemeint, die in einer Ecke, vor einer Wand oder vor einer Spitze stattfinden. Links, rechts und frei werden immer richtig erkannt)

Der neue Datensatz lieferte nun für alle Messungen Frontal (100%) als Klassifikation. Wird NaiveBayes benutzt, so liefern alle Messungen (Datensatz mit Bewegung, ohne, in Grips bereitgestellte Beispiel csv-Daten) eine Wahrscheinlichkeit von 0.33 für Ecke, Spitze, Frontal.

Werden die Beispieldaten aus Grips mit RandomForest kombiniert, so liefern alle Messung fast immer 60% für Spitze, 30% für Ecke und 10% für Frontal. Einzige Ausnahme ist für den Fall, wenn nur der vorderste Sensor Messwerte liefert und die anderen Sensoren „blind“ sind. Dieser wird als 100% Frontal klassifiziert.