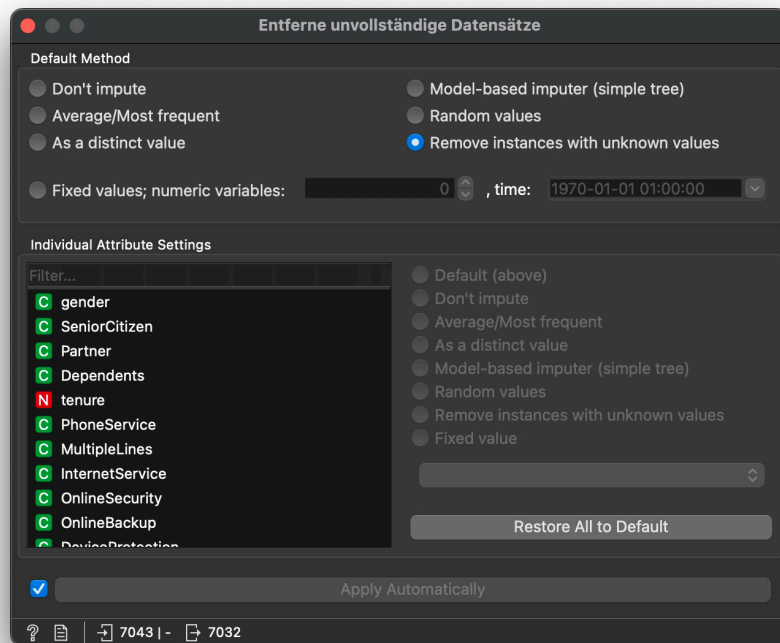


# Hinweise zu Orange Widgets für die Aufgaben 4, 5 und 6 in den Kapiteln 1.11 und 1.12



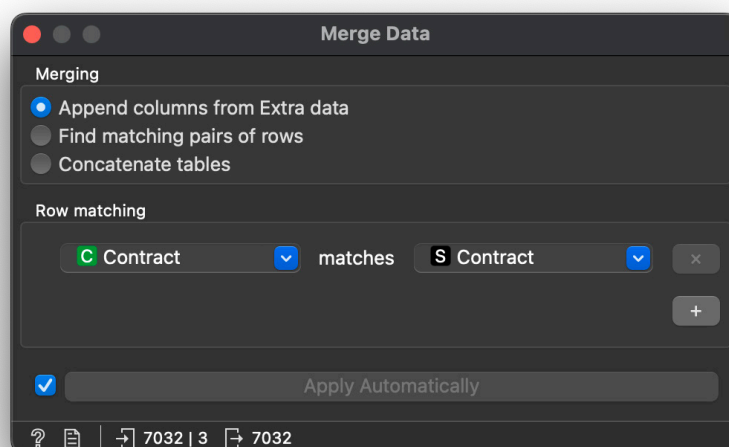
Impute

Kann dazu verwendet werden, Datensätze zu bereinigen (z.B. wie im Beispiel unten von unvollständigen Datensätzen).



Merge Data

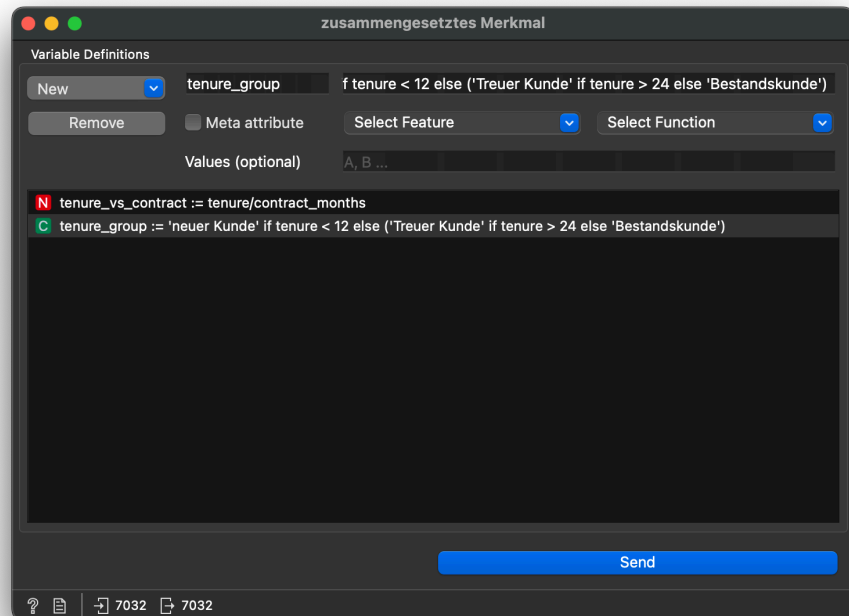
Kann dazu verwendet werden zwei Tabellen zu vereinigen (ähnlich zu einem SQL Join Befehl).





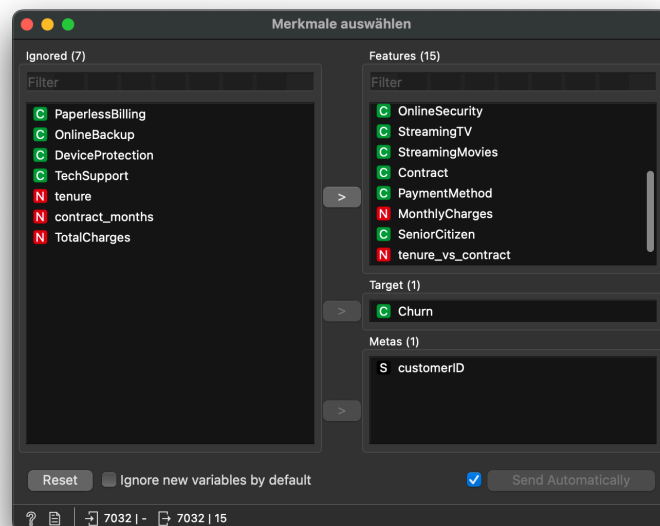
Formula






Ermöglicht das Erstellen neuer Merkmale basierend auf alten Merkmalen. Es können sowohl mathematisch Formeln (\*, /, +, -) als auch if then else Anweisungen verwendet werden (siehe Beispiel unten)

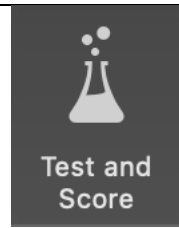


Select Columns

Hiermit lassen sich Merkmale auswählen, die für das Training des Modells herangezogen werden. Die zur Vorhersage verwendeten Merkmale finden sich im Feld „Features“, das vorhersagte steht unter „Target“. In Feld „Metas“ sind Informationen, die nicht zum Training verwendet werden sollen, aber als Metainformationen weiter mitgeführt werden soll.



 <p>Constant</p>	<p>Modell, das das häufigste Label bzw. den Durchschnittswert vorhersagt.</p>
 <p>Tree</p>	<p>Modell zur Klassifikation mit Entscheidungsbaum</p>
 <p>Random Forest</p>	<p>Modell zur Klassifikation mit Ensemble von Entscheidungsbäumen</p>
 <p>kNN</p>	<p>Modell zur Klassifikation anhand der k nächsten Nachbarn</p>
 <p>Neural Network</p>	<p>Kann dazu verwendet werden (mehrschichtige) neuronale Netze zu erzeugen. Im Feld „Neurons in hidden layers“ können die Neuronen auf den verschiedenen Schichten durch Kommas getrennt angegeben werden.</p> <div data-bbox="411 1200 1131 1906"> <p>Name</p> <p>Neural Network</p> <p>Neurons in hidden layers: 100,</p> <p>Activation: ReLu</p> <p>Solver: Adam</p> <p>Regularization, <math>\alpha=0.0001</math>:</p> <p>Maximal number of iterations: 200</p> <p><input checked="" type="checkbox"/> Replicable training</p> <p>Cancel <input checked="" type="checkbox"/> Apply Automatically</p> </div>



## Test and Score

Kann dazu verwendet werden, das Modell zu testen und verschiedene Metriken zu berechnen.

☐ Cross validation  
 Number of folds: 5  
☒ Stratified  
☐ Cross validation by feature  
☐ Random sampling  
 Repeat train/test: 10  
 Training set size: 66 %  
☒ Stratified  
☐ Leave one out  
☐ Test on train data  
☒ Test on test data

Evaluation results for target (None, show average over classes)
 

Model	AUC	CA	F1	Prec	Recall	MCC
kNN	0.797	0.780	0.775	0.772	0.780	0.421

Compare models by: Area under ROC curve  
 Negligible diff.: 0.1

	kNN
kNN	


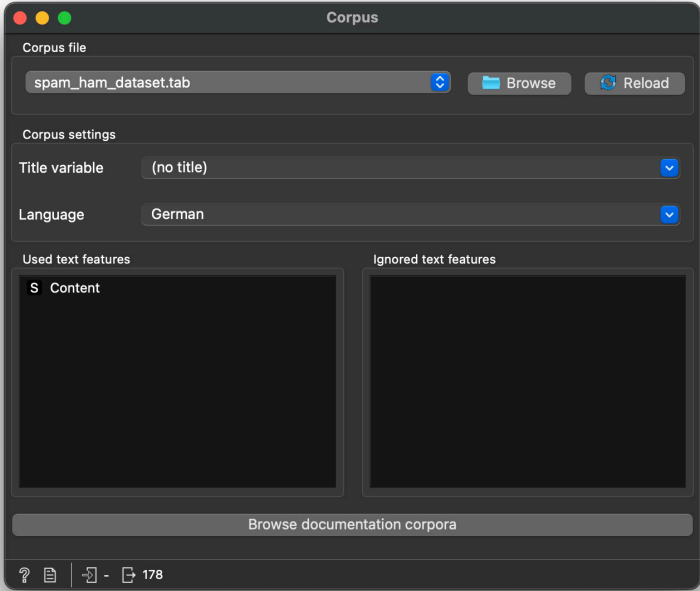

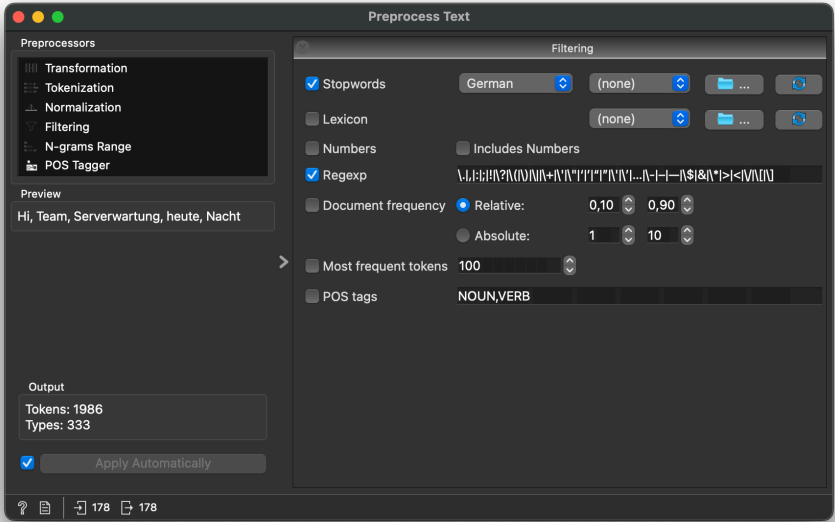

Table shows probabilities that the score for the model in the row is higher than that of the model in the column. Small numbers show the probability that the difference is negligible.

Dabei sollte darauf geachtet werden, dass rechts oben die „Evaluation results for target“ auf dem „average over classes“ stehen. Links lassen sich zudem verschiedene Strategien auswählen. Bei einem Test-/Trainingsdaten-Split sollte „Test on test data“ ausgewählt werden.

Damit das Applet seine Arbeit machen kann, sollte die Datenquelle zweimal verbunden werden.



Widgets für den Umgang mit Text als Daten (insbesondere Aufgabe 5)

 <p>Corpus</p>	<p>Wird verwendet um Textdaten zu importieren (.tab Datei steht für Aufgabe 5 bereit).</p> 
 <p>Preprocess Text</p>	<p>Wird verwendet, um Text vorzuverarbeiten. Im unteren Beispiel werden Sonderzeichen und deutsche Stopwörter entfernt. Es können aber andere Vorverarbeitungsschritte wie PoS Tagging (Tagging, ob es sich bei einem Wort um ein Verb etc handelt) angewendet werden.</p> 
 <p>Bag of Words</p>	<p>Wird verwendet um ein einfaches Bag-of-Words-Modell oder auch ein TF/IDF-Modell zu erzeugen.</p>

Hinweise zu Orange Widgets für die Aufgaben 4, 5 und 6 in den Kapiteln 1.11 und 1.12

Natürlich gibt es noch viele weitere Widgets, die sich für Ihre Projekte eignen, die obigen Tabellen stellt lediglich eine Auswahl dar, die geeignet sind, um die gegebenen Aufgaben zu bearbeiten, es sind jedoch auch viele andere Widgets einsetzbar.