

## Mögliche Eingaben zur Konfiguration der Multilabel-Klassifikation

„optional“ bedeutet, dass das Feld leer gelassen werden kann. Bitte unbedingt die Reihenfolge der Attribute beibehalten!

| Attribut                  | Wert   |
|---------------------------|--|
| <b>inputFile</b>          | Pfad zur .xlsx-Datei, in der die zu klassifizierenden Stellenanzeigen stehen   |
| <b>outputFolder</b>       | Pfad zum Ordner, in den die .xlsx-Datei mit den klassifizierten Dateien gelegt werden soll   |
| <b>focusesFile</b>        | Pfad zur .xlsx-Datei, in der die Schwerpunkte definiert sind   |
| <b>output</b>             | „ranking“: Ausgabe aller Schwerpunkte mit tatsächlichen Wahrscheinlichkeiten; „labels“: Liste mit Schwerpunkten, deren Wahrscheinlichkeiten über dem Schwellwert liegen                                    |
| <b>ignoreStopwords</b>    | „true“: Füllwörter werden nicht in den Merkmalsvektor aufgenommen; „false“: Merkmalsvektor wird mit Füllwörtern gebildet   |
| <b>normalizeInput</b>     | „true“: Tokens werden für den Merkmalsvektor normalisiert; „false“: Tokens werden nicht normalisiert   |
| <b>useStemmer</b>         | „true“: Tokens werden für den Merkmalsvektor gestemmt; „false“: Tokens werden nicht gestemmt   |
| <b>nGrams</b>             | Länge(n) der N-Gramme (als Ganzzahl, mit Kommata getrennt) (optional)  |
| <b>distance</b>           | „cosinus“, „euklid“ oder „manhattan“ als Distanzmaß im Vektorraum (nur, falls es sich um <i>ML-kNN</i> handelt)  |
| <b>threshold</b>          | Schwellwert, ab welcher Wahrscheinlichkeit ein Label gesetzt werden soll (als Double; min.: 0.0, max.: 1.0) (optional)   |
| <b>knnValue</b>           | Anzahl der betrachteten nächsten Nachbarn (als Ganzzahl) (nur, falls es sich um <i>ML-kNN</i> handelt)   |
| <b>allowemptylabelmap</b> | „true“: falls bei einer Stellenanzeige keine Label-Wahrscheinlichkeit den Schwellwert überschreitet, soll kein Label gesetzt werden; „false“: in o.g. Fall soll das wahrscheinlichste Label gesetzt werden |
| <b>classifier</b>         | „mlknn“ oder „naivebayes“ als Klassifikator  |
| <b>quantifier</b>         | „tfidf“ oder „loglikelihood“ als Art der Vektorgewichtung  |