A9)

Für die ersten Versuche verwenden wir nur einen Teil der Trainingsdaten, indem wir den Filter ... anwenden. Wir testen verschiedene Regelerner auf 5000 Trainingsdaten mit 10 fold cross –validation.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 10 fold cross-validation with 5000 DP | Jrip | NaiveBayes | IBK | ConjunctiveRule | DecisionTable | OneR | PART | J48 |
| Correctly Classified | 74,66 | 75,64 | 89,48 | 19,76 | 56,34 | 19,18 | 76,36 | 73,14 |
| TP Rate | 0,747 | 0,756 | 0,895 | 0,198 | 0,563 | 0,192 | 0,764 | 0,731 |
| FP Rate | 0,028 | 0,027 | 0,012 | 0,095 | 0,049 | 0,092 | 0,026 | 0,03 |
| Precision | 0,749 | 0,775 | 0,897 | 0,077 | 0,558 | 0,154 | 0,763 | 0,731 |

Die besten fünf Lerner testen wir erneut mit 20 fold cross-validation auf 10000 Trainingsdaten.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 20 fold cross-validation with 10000 DP | Jrip | NaiveBayes | IBK | PART | J48 |
| Correctly Classified | 78,82 | 76,87 | 90,93 | 80,76 | 76,78 |
| TP Rate | 0,788 | 0,769 | 0,909 | 0,808 | 0,768 |
| FP Rate | 0,024 | 0,026 | 0,01 | 0,021 | 0,026 |
| Precision | 0,789 | 0,783 | 0,911 | 0,808 | 0,768 |

Das beste Ergebnis liefert IBK, also nearest neighbour. Deshalb testen wir für diesen Algorithmus verschiedene N (Anzahl betrachteter Nachbarn). Auch hier betrachten wir wieder 20 fold cross-validation auf 10000 Trainingsdaten. Die Werte sind generell gut mit einer Precision Rate von über 0,9. Die besten Werte werden mit N = 5 erzielt.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| IBK-N Neighbours | N=1 | 2 | 4 | 5 | 6 | 7 | 9 | 10 | 100 |
| Correctly Classified | 90,93 | 91,39 | 93,84 | 94,25 | 94,01 | 94,04 | 93,85 | 93,76 | 87,9 |
| TP Rate | 0,909 | 0,914 | 0,938 | 0,943 | 0,94 | 0,94 | 0,939 | 0,938 | 0,879 |
| FP Rate | 0,01 | 0,01 | 0,007 | 0,007 | 0,007 | 0,007 | 0,007 | 0,007 | 0,014 |
| Precision | 0,911 | 0,918 | 0,94 | 0,944 | 0,942 | 0,942 | 0,94 | 0,94 | 0,893 |

Wir wollen die Testdaten mit IBK-5 klassifizieren. Unsere Vorhersage mit 10 fold cross-validation auf dem kompletten Trainingsdatensatz besagt:

|  |  |
| --- | --- |
|  | IBK - 5 |
| Correctly Classified | 96,598 |
| TP Rate | 0,966 |
| FP Rate | 0,004 |
| Precision | 0,966 |