15

Table, calendar

Description automatically generated

16

Full tree: Not useful at all. Too much noise. Minimum class sizes and maximum number of steps: Gets rid of the 'noise' of a full tree. Useful for tactical decision making. Also useful for operational businesses when you need to make quick decisions, these are practical compared to the full tree. Benefits: easy to interpret, versatile, low cost to produce Cons: overfitting and not complete when it comes to complexer problems.

17a & b

Text

Description automatically generated with medium confidence

17c

-2.5444 \* (normalized) satisfaction + 0.2106 \* (normalized) evaluation + -0.246 \* (normalized) projects + 0.6457 \* (normalized) hours + 0.7506 \* (normalized) tenure + -0.9469 \* (normalized) accident=yes + -0.208 \* (normalized) promotion=yes + -0.003 \* (normalized) department=technical + -0.1445 \* (normalized) department=mktg\_sales + 0.0927 \* (normalized) department=management + 0.0455 \* (normalized) department=product\_rand + 0.079 \* (normalized) department=acct\_it + -0.0697 \* (normalized) department=support\_hr + 0.5104 \* (normalized) salary=low + -0.7065 \* (normalized) salary=high + 0.1961 \* (normalized) salary=medium - 0.3215

18

A picture containing table

Description automatically generated

19 a

|  |  |
| --- | --- |
| Method | Correctly classified |
| No analysis | 76,1000% |
| Logistic Regression | 78,4154% |
| Decision Stump | 82,0692% |
| Full classification tree | 100,0000% |
| Tree limited by leaf size | 95,3923% |
| Tree limited by path length | 95,4923% |
| SVM Linear | 77,8846% |
| SVM Curvilinear | 93,7692% |

B

The limited tree’s together with the SVM Curvilinear are the closest to perfect in correctly classifying instances. You could argue for the decision stump, but the others do not significantly differ from no analysis. A high percentage of correctly classified instances is a useful indicator, but it doesn’t necessarily mean that you have a good model. There are other statistics like the Kappa statistic and the detailed accuracy by class statistics, like precision, recall, F-measure and ROC area and finally the confusion matrix.