# Dataset for M.L. training

Filename : [parkinsons\_updrs.data](https://archive.ics.uci.edu/ml/machine-learning-databases/parkinsons/telemonitoring/parkinsons_updrs.data)

Link: https://archive.ics.uci.edu/ml/machine-learning-databases/parkinsons/telemonitoring/

Parkinsons Telemonitoring Data Set

Abstract: Oxford Parkinson's Disease Telemonitoring Dataset

============================================================

Data Set Characteristics: Multivariate

Attribute Characteristics: Integer, Real

Associated Tasks: Regression

Number of Instances: 5875

Number of Attributes: 26

Area: Life

Date Donated: 2009-10-29

# Characteristics of used variables from dataset before normalization.

HNR

|  |  |
| --- | --- |
| Minimum | 1.659 |
| Maximum | 37.875 |
| Mean | 21.679 |
| StdDev | 4.291 |

RPDE

|  |  |
| --- | --- |
| Minimum | 0.151 |
| Maximum | 0.966 |
| Mean | 0.541 |
| StdDev | 0.101 |

DFA

|  |  |
| --- | --- |
| Minimum | 0.514 |
| Maximum | 0.866 |
| Mean | 0.653 |
| StdDev | 0.071 |

PPE

|  |  |
| --- | --- |
| Minimum | 0.022 |
| Maximum | 0.732 |
| Mean | 0.220 |
| StdDev | 0.091 |

Dataset Before normalization some values:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| subject# | HNR | RPDE | DFA | PPE | total\_UPDRS |
| 1 | 21.64 | 0.41888 | 0.54842 | 0.16006 | 34.398 |
| 2 | 27.183 | 0.43493 | 0.56477 | 0.1081 | 34.894 |
| 3 | 23.047 | 0.46222 | 0.54405 | 0.21014 | 35.389 |
| 4 | 24.445 | 0.4873 | 0.57794 | 0.33277 | 35.81 |
| 5 | 26.126 | 0.47188 | 0.56122 | 0.19361 | 36.375 |
| 6 | 22.946 | 0.53949 | 0.57243 | 0.195 | 36.87 |
| 7 | 22.506 | 0.4925 | 0.54779 | 0.17563 | 37.363 |
| 8 | 22.929 | 0.47712 | 0.54234 | 0.23844 | 37.857 |
| 9 | 22.078 | 0.51563 | 0.61864 | 0.20037 | 38.353 |
| 10 | 22.606 | 0.50032 | 0.58673 | 0.20117 | 38.849 |
|  |  |  |  |  |  |

# Variables characteristics after normalization

The normalization processes was made using the WEKA software , following the process:

Normalized Value = [original value – minimum value] / [maximum value – minimum value]

* minimum and maximum value for dataset

HNR

|  |  |
| --- | --- |
| Minimum | 0 |
| Maximum | 1 |
| Mean | 0.553 |
| StdDev | 0.118 |

RPDE

|  |  |
| --- | --- |
| Minimum | 0 |
| Maximum | 1 |
| Mean | 0.479 |
| StdDev | 0.124 |

DFA

|  |  |
| --- | --- |
| Minimum | 0 |
| Maximum | 1 |
| Mean | 0.396 |
| StdDev | 0.202 |

PPE

|  |  |
| --- | --- |
| Minimum | 0 |
| Maximum | 1 |
| Mean | 0.278 |
| StdDev | 0.129 |

Dataset after normalization, some values:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| subject# | HNR | RPDE | DFA | PPE | total\_UPDRS |
| 1 | 0.551717 | 0.328638 | 0.097793 | 0.194544 | 34.398 |
| 2 | 0.704771 | 0.34833 | 0.1443 | 0.121335 | 34.894 |
| 3 | 0.590568 | 0.381812 | 0.085362 | 0.265104 | 35.389 |
| 4 | 0.629169 | 0.412583 | 0.181761 | 0.437884 | 35.81 |
| 5 | 0.675585 | 0.393664 | 0.134202 | 0.241814 | 36.375 |
| 6 | 0.587779 | 0.476615 | 0.166088 | 0.243773 | 36.87 |
| 7 | 0.57563 | 0.418963 | 0.096001 | 0.216481 | 37.363 |
| 8 | 0.587309 | 0.400093 | 0.080498 | 0.304978 | 37.857 |
| 9 | 0.563812 | 0.447341 | 0.297531 | 0.251339 | 38.353 |
| 10 | 0.578391 | 0.428557 | 0.206764 | 0.252466 | 38.849 |

# Splitting total UPDRS in 10 groups representing a generic grade.

The Total UPDRS is not useful for users , for this reason this grade has been translated in a generic grade like 1 to 10 for represent the gravity of disease.

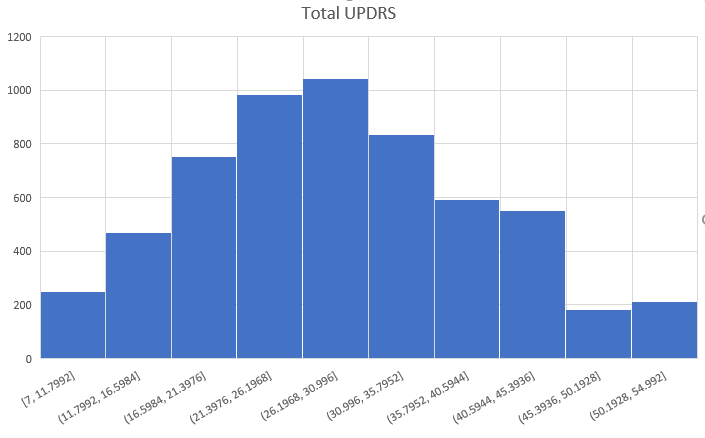


Figure 1- Dataset divided in 10 groups of Parkinson

|  |  |
| --- | --- |
| Generic Grade | Total UPDRS interval |
| 1 | 7 to 11.8 |
| 2 | 11.8 to 16.6 |
| 3 | 16.6 to 21.40 |
| 4 | 21.40 to 26.2 |
| 5 | 26.2 to 31 |
| 6 | 31 to 35.80 |
| 7 | 35.80 to 40.6 |
| 8 | 40.6 to 45.40 |
| 9 | 45.40 to 50.20 |
| 10 | 50.20 to 55 |

Dataset after normalization and translation of UPDRS to Generic grade

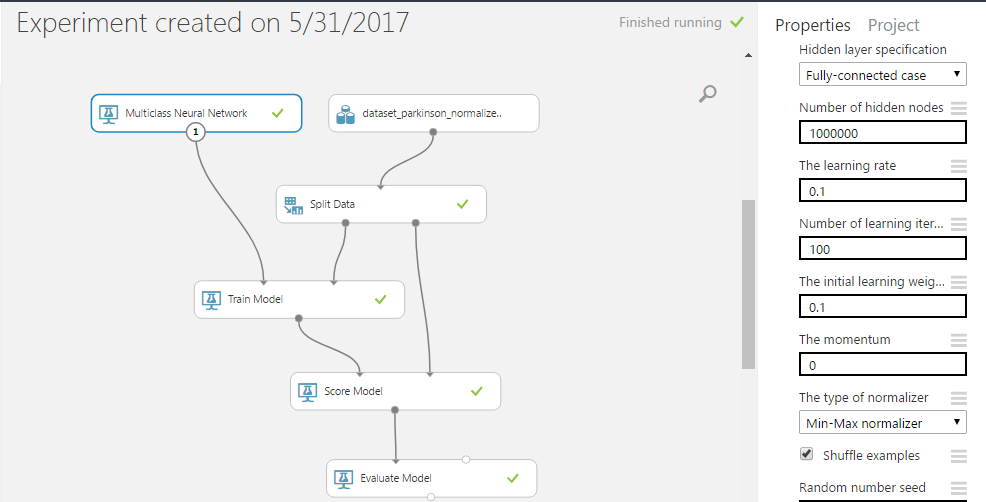
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| subject# | HNR | RPDE | DFA | PPE | total\_UPDRS |
| 1 | 0.551717 | 0.328638 | 0.097793 | 0.194544 | 6 |
| 2 | 0.704771 | 0.34833 | 0.1443 | 0.121335 | 6 |
| 3 | 0.590568 | 0.381812 | 0.085362 | 0.265104 | 6 |
| 4 | 0.629169 | 0.412583 | 0.181761 | 0.437884 | 7 |
| 5 | 0.675585 | 0.393664 | 0.134202 | 0.241814 | 7 |
| 6 | 0.587779 | 0.476615 | 0.166088 | 0.243773 | 7 |
| 7 | 0.57563 | 0.418963 | 0.096001 | 0.216481 | 7 |
| 8 | 0.587309 | 0.400093 | 0.080498 | 0.304978 | 7 |
| 9 | 0.563812 | 0.447341 | 0.297531 | 0.251339 | 7 |
| 10 | 0.578391 | 0.428557 | 0.206764 | 0.252466 | 7 |
|  |  |  |  |  |  |

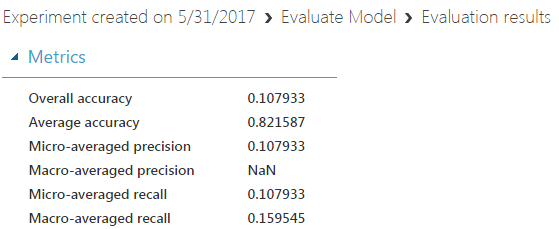
# Machine learning creation

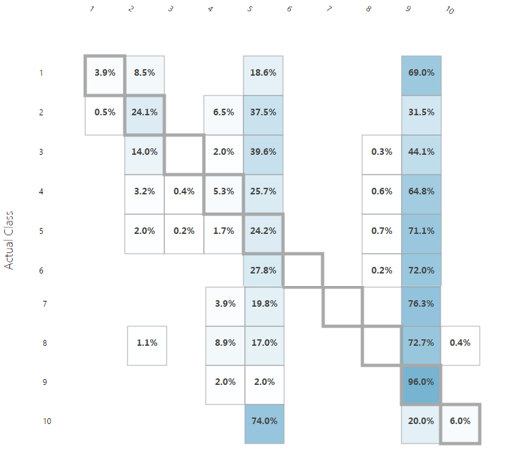
For machine learning classification we use a Online service called Microsoft Azure. This is a cloud computing platform that includes a machine learning ambient called Machine learning studio.

Fist process is a training using a dataset discussed before.

**Test with neural networks**







**Test with Multiclass Decision Jungle**

module to create a machine learning model that is based on a supervised learning algorithm called *decision jungles*. The model can be used to predict a target that has multiple values. The module returns an untrained classifier that can be passed to another module, such as [Train Model](https://msdn.microsoft.com/en-us/library/azure/dn906044.aspx) to [Tune Model Hyperparameters](https://msdn.microsoft.com/en-us/library/azure/dn905810.aspx), for training on a labeled training data set. The trained model can then be used to make predictions. Alternatively, the untrained model can be passed to [Cross-Validate Model](https://msdn.microsoft.com/en-us/library/azure/dn905852.aspx) for cross-validation against a labeled data set.

