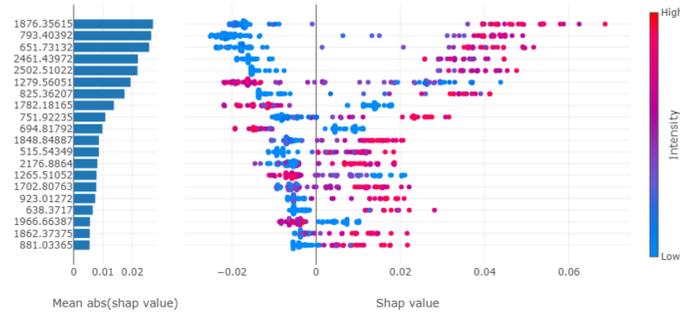


How to perform Shapley values in our platform?

CLOVER MS Data Analysis Software

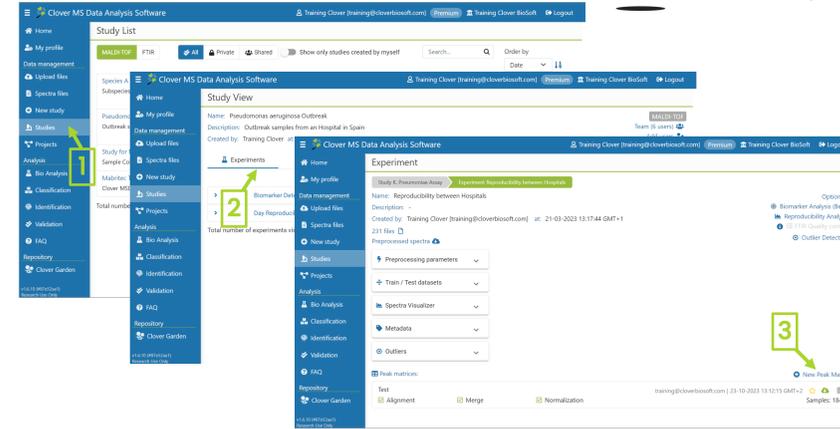
What is it?



In Clover Biosoft, it is possible to perform Shapley values in a Random Forest analysis. In our case, **Shapley values** measure the feature importance, based on its average marginal contribution considering all possible combinations, in order to establish a classification between different categories.

1 First steps

This guide is the same for MALDI and FTIR spectra

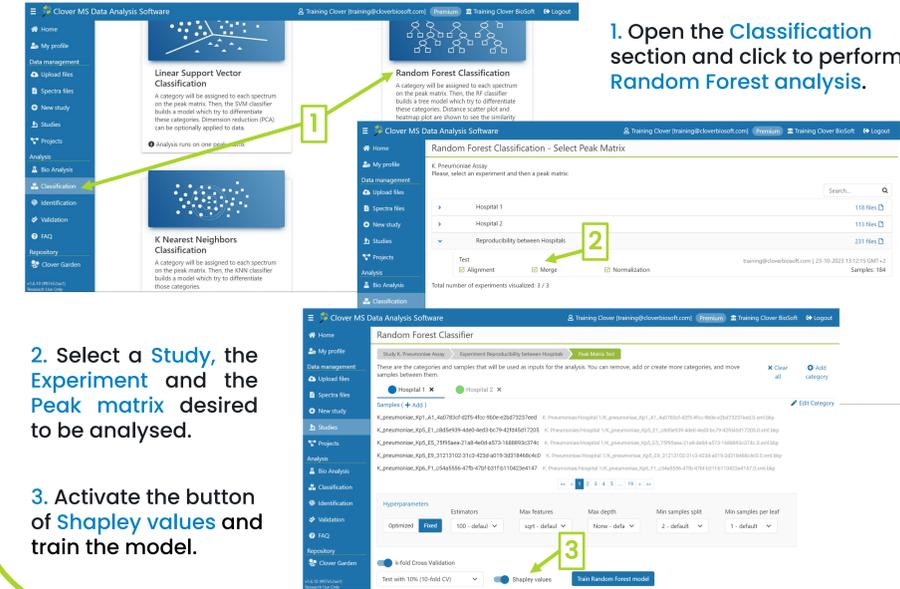


Create and/or choose a Study

Create and/or choose an Experiment

Create a Peak Matrix

2 Random Forest



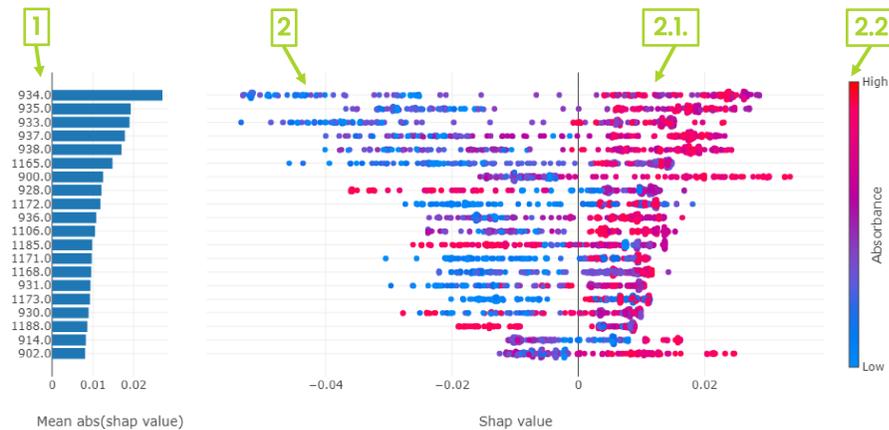
1. Open the **Classification** section and click to perform **Random Forest analysis**.

2. Select a **Study**, the **Experiment** and the **Peak matrix** desired to be analysed.

3. Activate the button of **Shapley values** and train the model.

3 Shapley values

The results from Shapley values are the following:



1. A list of the main features that contribute (to have a function) in the classification: **the higher** the feature position in the list is, **the more important** it is to discriminate your picked category.

2. A graph in which **every sample is represented by a point** for each feature. This plot gives us different information:

2.1. **The more to the right** the point is in the plot, the **highest contribution** the point has in the picked category. However, if the point is on the left side, its function contributes to the opposite category.

2.2. The **colour** indicates if the feature has a high value (in red) or a low value (in blue).

In this example, the first five features (934, 935, 933, 937, and 938) have a high contribution in the discrimination of the picked category. Indeed, these features contribute to the positive class identification.



cloverbiosoft.com
clovermsdataanalysis.com

Clover Bioanalytical Software
Copyright © 2015

Edif. Centro de Empresas PTS.
Av. del Conocimiento, 41
18016 Granada, Spain

Phone: +34 958 991 543
email: info@cloverbiosoft.com
support@cloverbiosoft.com

Clover MS Data Analysis Software, Quick Start Guide



What is it?



In **Clover Biosoft**, it is possible to perform Shapley values in a Random Forest analysis. In our case, **Shapley values** measure the feature importance, based on its average marginal contribution considering all possible combinations, in order to establish a classification between different categories.



cloverbiosoft.com
clovermsdataanalysis.com



Clover Bioanalytical Software
Copyright © 2015

Edif. Centro de Empresas PTS.
Av. del Conocimiento, 41
18016 Granada, Spain

Phone: +34 958 991 543
email: info@cloverbiosoft.com
support@cloverbiosoft.com

Clover MS Data Analysis Software, Quick Start Guide

How to perform Shapley values in our platform?

CLOVER MS Data Analysis Software

1

First steps



This guide is the same for MALDI and FTIR spectra

The screenshot displays the software's user interface. The top navigation bar includes the user's name 'Training Clover', email 'training@cloverbiosoft.com', and account type 'Premium'. The left sidebar contains a menu with categories like 'Data management', 'Analysis', and 'Repository'. The main content area is divided into three overlapping panels: 'Study List', 'Study View', and 'Experiment'. The 'Study List' panel shows a table of studies with filters for 'MALDI-TOF' and 'FTIR'. The 'Study View' panel shows details for a study named 'Pseudomonas aeruginosa Outbreak'. The 'Experiment' panel shows details for an experiment named 'Reproducibility between Hospitals' and includes a 'Peak matrices' section with checkboxes for 'Alignment', 'Merge', and 'Normalization'. A 'New Peak Matrix' button is visible in the bottom right of the 'Experiment' panel.

Create and/or
choose a **Study**

Create and/or
choose an
Experiment

Create a
Peak Matrix



cloverbiosoft.com
clovermsdataanalysis.com

Clover Bioanalytical Software
Copyright © 2015

Edif. Centro de Empresas PTS.
Av. del Conocimiento, 41
18016 Granada, Spain

Phone: +34 958 991 543
email: info@cloverbiosoft.com
support@cloverbiosoft.com

Clover MS Data Analysis Software, Quick Start Guide



How to perform Shapley values in our platform?

CLOVER MS Data Analysis Software

2

Random Forest

The figure consists of three screenshots of the Clover MS Data Analysis Software interface, illustrating the steps to perform a Random Forest analysis with Shapley values.

Step 1: The first screenshot shows the 'Classification' section of the software. The 'Random Forest Classification' option is highlighted with a green arrow and a '1' in a box. A text box on the right says: "1. Open the Classification section and click to perform Random Forest analysis."

Step 2: The second screenshot shows the 'Random Forest Classification - Select Peak Matrix' dialog. The 'Reproducibility between Hospitals' experiment is selected, and the 'Merge' button is highlighted with a green arrow and a '2' in a box. A text box on the left says: "2. Select a Study, the Experiment, and the Peak matrix desired to be analysed."

Step 3: The third screenshot shows the 'Random Forest Classifier' configuration page. The 'Shapley values' checkbox is checked and highlighted with a green arrow and a '3' in a box. A text box on the left says: "3. Activate the button of Shapley values and train the model."

2. Select a Study, the Experiment, and the Peak matrix desired to be analysed.

3. Activate the button of Shapley values and train the model.



cloverbiosoft.com
clovermsdataanalysis.com



Clover Bioanalytical Software
Copyright © 2015

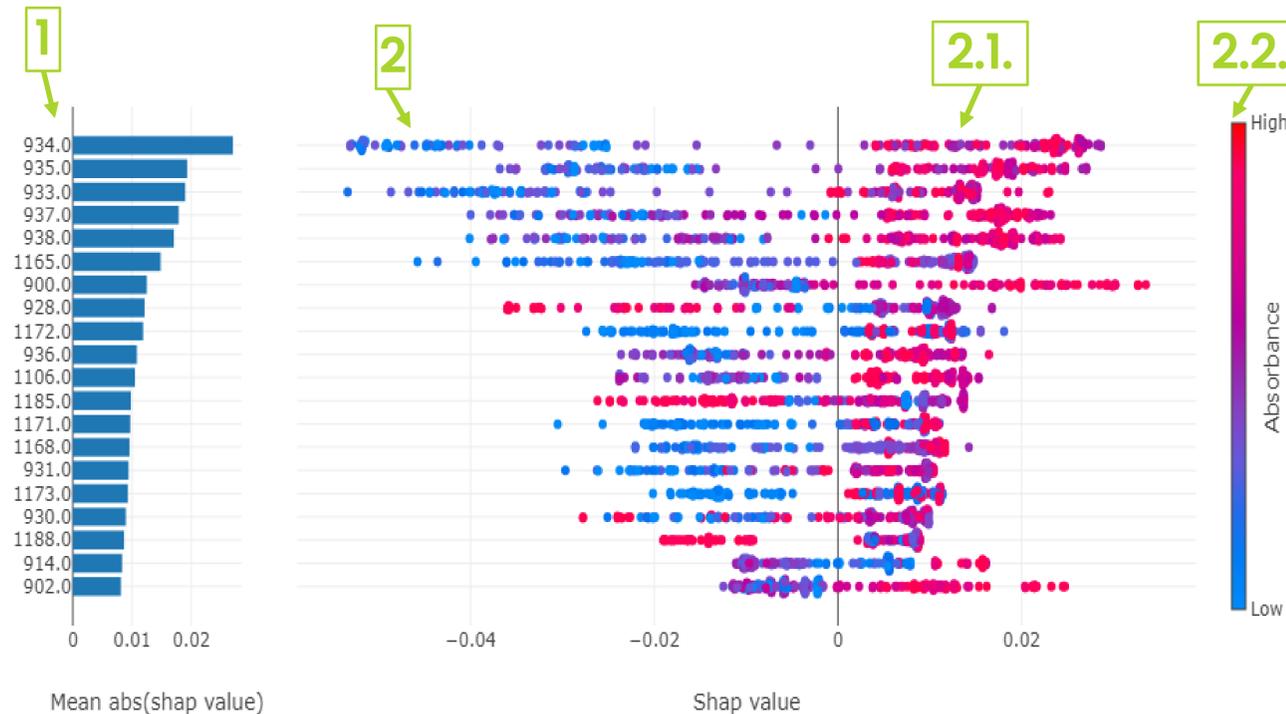
Edif. Centro de Empresas PTS.
Av. del Conocimiento, 41
18016 Granada, Spain

Phone: +34 958 991 543
email: info@cloverbiosoft.com
support@cloverbiosoft.com

Clover MS Data Analysis Software, Quick Start Guide

3 Shapley values

The results from Shapley values are the following:



1. A list of the main features that contribute (to have a function) in the classification: the higher the feature position in the list is, the more important it is to discriminate your picked category.

2. A graph in which every sample is represented by a point for each feature. This plot gives us different information:

2.1. The more to the right the point is in the plot, the highest contribution the point has in the picked category. However, if the point is on the left side, its function contributes to the opposite category.

2.2. The colour indicates if the feature has a high value (in red) or a low value (in blue).

In this example, the first five features (934, 935, 933, 937, and 938) have a high contribution in the discrimination of the picked category. Indeed, these features contribute to the positive class identification.



cloverbiosoft.com
clovermsdataanalysis.com



Clover Bioanalytical Software
Copyright © 2015

Edif. Centro de Empresas PTS.
Av. del Conocimiento, 41
18016 Granada, Spain

Phone: +34 958 991 543
email: info@cloverbiosoft.com
support@cloverbiosoft.com

Clover MS Data Analysis Software, Quick Start Guide