

# Scio-machine.com

Machine learning software  
easy to use and running in a browser.

## Quick start 1.0

### Example 1

An agri-food company wants to develop and market a variety of peas with low sugar content while maintaining a pleasant flavor.

What characteristics should this variety have?

Working data: Evaluation by a jury of the taste criteria of several varieties of peas.

### Example 2

In order to optimize its financial operations a company wants to know in advance the evolution of the Australian dollar AUD versus the New Zealand dollar NZD.

What is the 3-day trend for AUD/NZD?

Working data: quotation history of currency pairs on the exchange market.

### Features & Use

General operation.

Preparation of the data.

Parameters

That is to say.

Security and privacy.

## Example 1

### Background

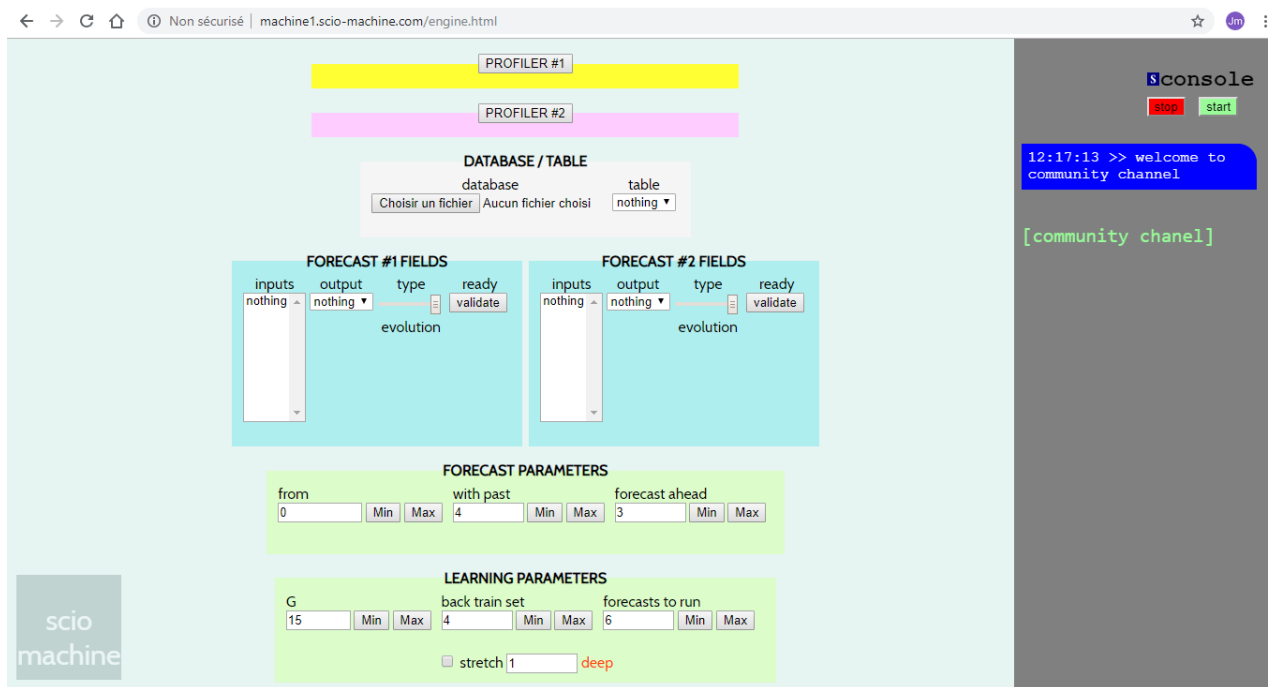
An agri-food company wants to develop and market a variety of peas with low sugar content while maintaining a pleasant flavor.

What characteristics should this variety have?

Working data: Evaluation by a jury of the taste criteria of several varieties of peas.

### The software

► Open the web page <http://machine1.scio-machine.com/engine.html> from a recent Chrome or Firefox browser.



### The data

You must be in possession of the SCIO-MACHINE-DEMO-DATABASE.sqlite database. It can be downloaded from the Help section.

► Load the SCIO-MACHINE-DEMO-DATABASE.sqlite database.

► Select the peas table.

Information on the peas table at <http://openmv.net/info/peas> . You can browse the data in this table by opening the database from a manager such as DB Browser for SQLite <https://sqlitebrowser.org>

## Settings

► Configure the interface settings as follows:

**inputs:** FORECAST #1 and #2:

["Tenderometer", "Drymatter", "Drymatterafterfreezing", "SucrosePercent", "TotalGlucose1", "TotalGlucose2"]

They correspond to physico-chemical properties of the evaluated varieties.

**output** FORECAST #1: Flavour

**output** FORECAST #2: Sweet

Flavour and Sweet are criteria of taste evaluated by the jury.

The analysis will aim to produce 2 different forecasts from the same input fields.

**type** #1 and #2 = classification

Concerns the choice of the method of calculation of the forecasts. See the parameters table.

**validate** to confirm selections.

**from** = 8

Place the cursor on the 9<sup>th</sup> record of the table. Numbering starts at 0.

**with past** = 1

Specifies that only the record pointed by the cursor *from* will provide input values to calculate the output forecasts.

**forecast ahead** = 0

Indicates that the output *forecast* (Flavor or Sweet) will be made for the record pointed by the cursor *from*.

**G** = 60

Specifies the number of generations that the genetic selection algorithm will perform at each step of the learning process.

**back train set** = 9

Indicates a set of tests to be performed at each step from the records preceding the cursor *from*.

**Forecasts to run** = 1

Indicates the number of steps (the cursor *from* moves forward one position per step).

**Stretch** not checked

Means that the *back train set* follows the cursor without increasing its size.

**deep** = 2

Specifies the degree of complexity of the two neural networks.

**real plots to display** = 30

To display the next 30 real values of the two output, from the initial location of the cursor *from*.

**0/1 view** = no

Real data and forecasts will be displayed based on their true value.

**To do** = learn & forecast

Neural networks are reset, a step-by-step learning and forecasting process is performed.

## Learning

- ▶ Press the start button in the console.
- ▶ Check and confirm the configuration.

```
running mode:

database : SCIO-MACHINE-DEMO-
DATABASE.sqlite
table : peas
from : 8
forecast ahead : 0
with past : 1

G : 60
deep : 2
back train set : 9
stretch : false
forecasts to run : 1
real plots : 30
run todo : learn & forecast

view : real range

#1 field output : ["Flavour"]
#1 fields input : 6 fields
#1 type : classification

#2 field output : ["Sweet"]
#2 fields input : 6 fields
#2 type : classification

back to settings ok
```

The steps of the learning process are displayed in the console. The graph displays the actual values of the two output outputs (Flavour and Sweet) as well as the computed values which should not be interpreted at this stage as forecasts.

## Forecasting

Once the learning & forecast process has been completed, proceed to a forecast.

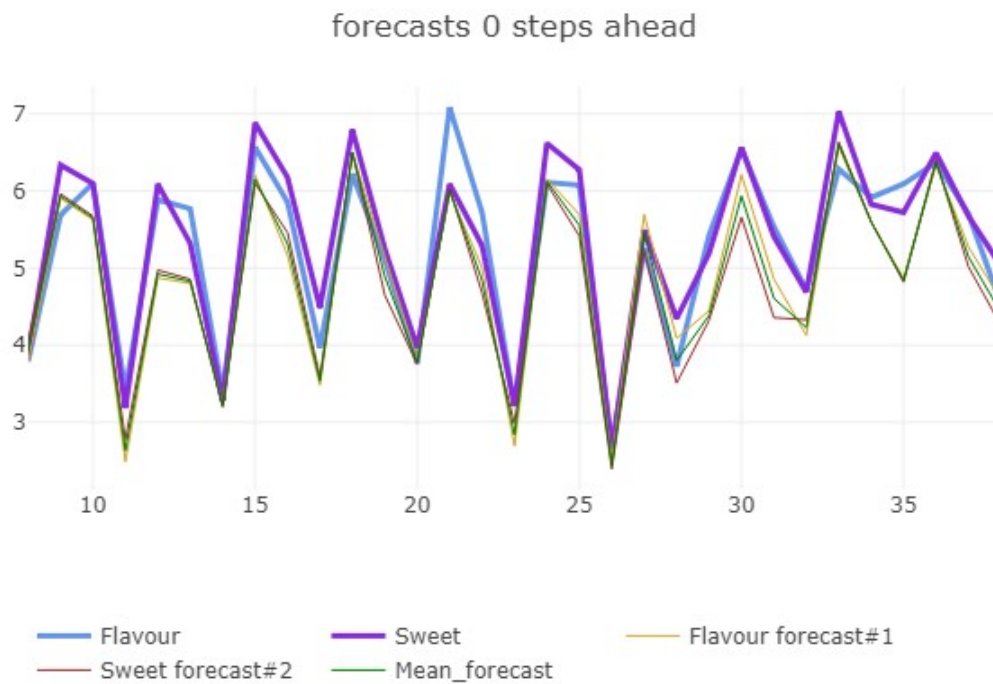
- ▶ Change the settings as follows:

**Forecasts to run:** 30

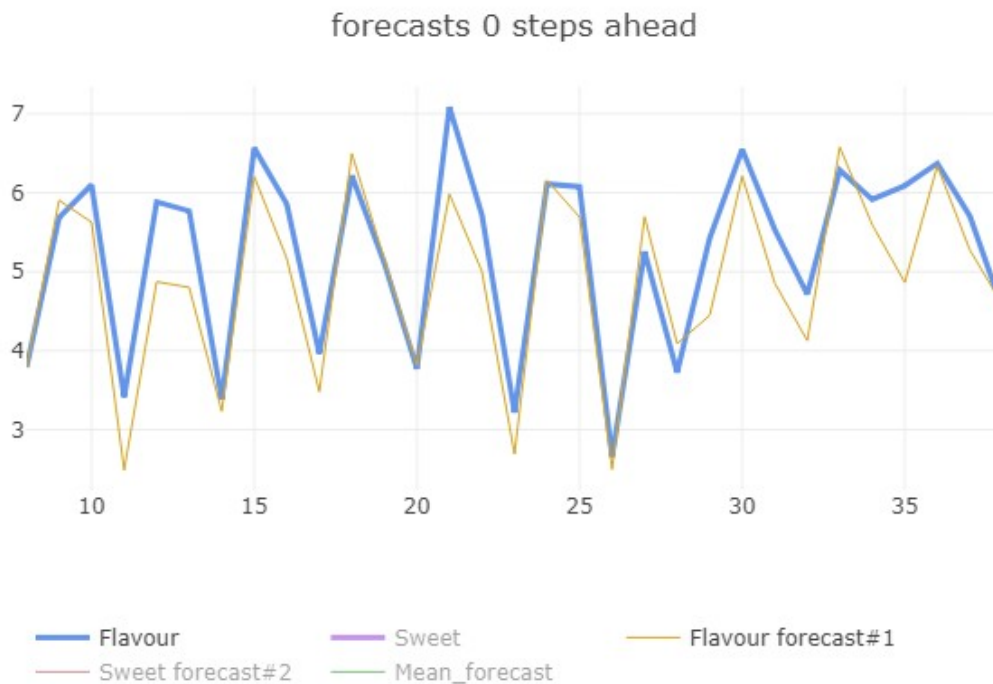
**to do:** forecast only

- ▶ Click *start* in the console.
- ▶ Check and accept the settings.

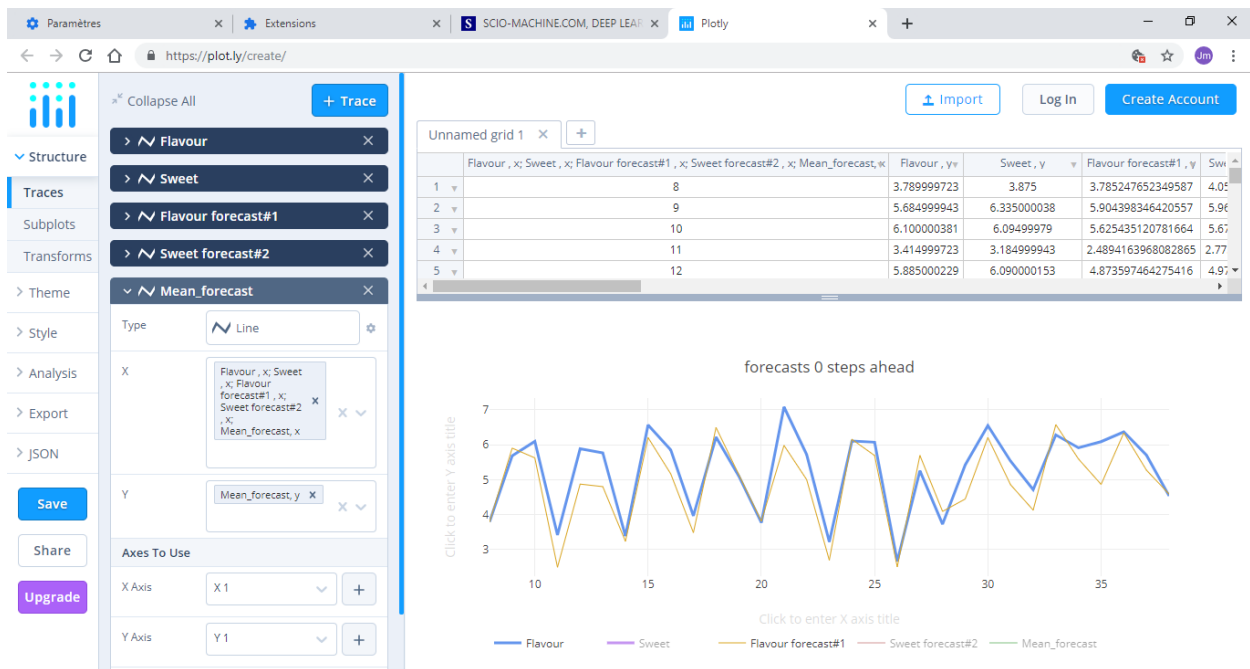
A process of calculation of forecasts begins. It uses the acquired state of both neural networks. The result obtained at the end of the process may be different from the graph below, this is due to the very character of the genetic selection algorithm. However it must have strong similarities.



It is noted on the one hand that the real values (in blue and light blue) of Flavour and Sweet are highly correlated, on the other hand that the forecasts obtained beyond record 9 are relevant. Lines can be hidden/displayed to appreciate results.



Note that all graphics produced can be saved independently on the plot.ly site by clicking on the floppy disk icon at the top of the graph.



## Report

You can print a report to keep track of forecasts and settings.

- ▶ Press *save report*.

## Profiling

Profiling allows to obtain a personalized forecast of the output by adjusting the input values. The last state of the neural network will be used.

Goal: Determine the best profile with the lowest sugars content and with a flavour (Flavour) rated 5 or more.

- ▶ Press PROFILER #1

- ▶ Adjust the input values

First adjust the input fields ["SucrosePercent", "TotalGlucose1", "TotalGlucose2"] to their minimum value in the profiler window. These three fields are the characteristics of sugar content. Then adjust the other input fields so that Flavour has the highest possible value. Compromises may be necessary.

- ▶ Comment the fields according to their impact on the Flavour value to keep track of them.

- ▶ Press *Save CSV*

The profile is saved. You can produce multiple profiles. Note that the neural network used to generate the profile is identified by a number. You can find this number in the report.

|    | A  | B | C                      | D       | E       | F      | G                                      |
|----|--|---|------------------------|---------|---------|--------|--|
| 1  | Profiler #1 - scio-machine.com                 |   |                        |         |         |        |  |
| 2  | generated date-time: 04/04/2019 à 22:01:51     |   |                        |         |         |        |  |
| 3  | generated network id: a2e187d2                 |   |                        |         |         |        |  |
| 4  | from: SCIO-MACHINE-DEMO-DATABASE.sqlite / peas |   |                        |         |         |        |  |
| 5  |  |   |                        |         |         |        |  |
| 6  | target:  |   |                        |         |         |        |  |
| 7  | Flavour 5.6381                                 |   |                        |         |         |        |  |
| 8  |  |   | inputs                 | minimum | maximum | value  | comment                                |
| 9  |  |   | Tenderometer           | 88,00   | 200,00  | 105,92 | very strong impact                     |
| 10 |  |   | Drymatter              | 13,80   | 28,10   | 20,66  | little impact                          |
| 11 |  |   | Drymatterafterfreezing | 17,91   | 28,75   | 18,56  | little impact                          |
| 12 |  |   | SucrosePercent         | 1,00    | 6,00    | 1,00   | should be set to lowest value          |
| 13 |  |   | TotalGlucose1          | 2,10    | 6,70    | 2,74   | little impact. Set this parameter low. |
| 14 |  |   | TotalGlucose2          | 2,20    | 7,00    | 2,49   | little impact. Set this parameter low. |
| 15 |  |   |                        |         |         |        |  |
| 16 |  |   |                        |         |         |        |  |

The file contains the values of the physico-chemical characteristics that the variety should have to meet the marketing requirements. The minimum and maximum values correspond to the respective values of the input field encountered in the table.

The file can be opened and formatted from a spreadsheet software.

This example has been created to show how Scio-machine.com works in the context of profile forecasting. Any forecast carries a notion of uncertainty. It is important to cross-reference forecasts with other assessments before making responsible decisions.

Mechanized prediction must be used with a benevolent ethic towards the human being and his interests.

## Example 2

### Background

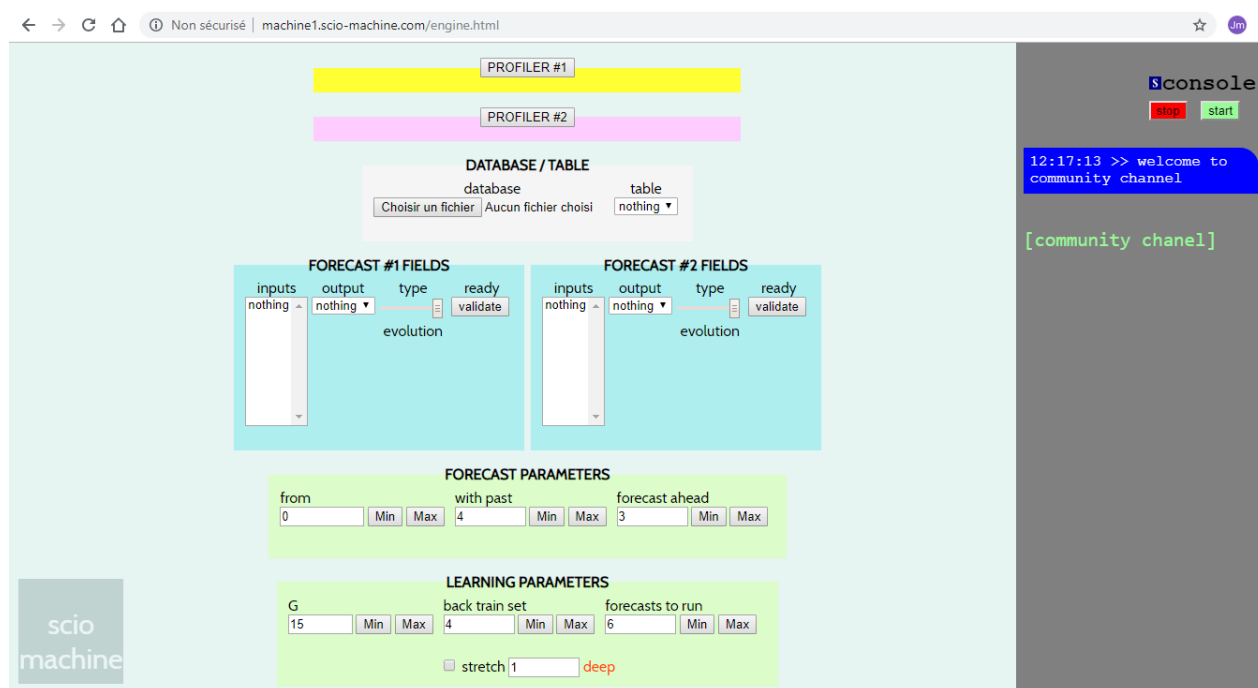
In order to optimize its financial operations a company wants to know in advance the evolution of the Australian dollar AUD versus the New Zealand dollar NZD.

What is the 3-day trend for AUD/NZD?

Working data: quotation history of currency pairs on the exchange market.

### The software

► Open the web page <http://machine1.scio-machine.com/engine.html> from a recent Chrome or Firefox browser.



### The data

You must be in possession of the SCIO-MACHINE-DEMO-DATABASE.sqlite database. It can be downloaded from the Help section.

► Load the SCIO-MACHINE-DEMO-DATABASE.sqlite database.

► Select the forex table.

The forex table was built from price history of several currency pairs. The data come from [https://www.dukascopy.com/plugins/fxMarketWatch/?historical\\_data](https://www.dukascopy.com/plugins/fxMarketWatch/?historical_data).

You can browse the data of this table by opening the database from a manager such as DB Browser for SQLite (<https://sqlitebrowser.org/>).



## Settings

► Configure the interface settings as follows:

**inputs:** FORECAST #1 and #2:

[AUDCADHigh, AUDCADLow, AUDCADClose, AUDCHFHigh, AUDCHFLow, AUDCHFClose, AUDNZDHigh, AUDNZDLow, AUDNZDClose, CADCHF High, CADCHFLow, CADCHFClose, NZDCADHigh, NZDCADLow, NZDCA Dclose, NZDCHFHigh, NZDCHFLow, NZDCHFClose]  
They correspond to the daily quotations High, Low, Close of the 6 pairs AUD/CAD, AUD/CHF, AUD/NZD, CAD/CHF, NZD/CAD and NZDCHF.

**output** FORECAST #1: AUDNZDClose

**output** FORECAST #2: AUDNZDClose

AUDNZDClose is the expected value.

Forecasts #1 and #2 will be based on the same AUDNZDClose output field from the same input fields. However, the results of the two forecasts may differ slightly, due to the fact that the genetic algorithm is an undetermined combinatorial process.

**type** #1 and #2= evolution

Concerns the choice of the method of calculation of the forecasts. See the parameters table.

**validate** to confirm selections.

**from** = 74

Place the cursor on the 75<sup>th</sup> record of the table. Numbering starts at 0.

**with past** = 4

Specifies that the record pointed by the cursor *from* and the three preceding it will provide input values to calculate the output forecasts.

**forecast ahead** = 3

Indicates that the output forecast (AUDNZDClose) will be made for the third record located to the right of the cursor *from*.

**G** = 10

Specifies the number of generations that the genetic selection algorithm will perform at each step of the learning process.

**back train set** = 2

Indicates a set of tests to be performed at each step from the records preceding the cursor *from*.

**forecasts to run** = 12

Indicates the number of steps (the cursor *from* moves forward one position per step).

**Stretch** not checked

Means that the *back train set* follows the cursor without increasing its size.

**deep** = 100

Specifies the degree of complexity of the two neural networks.

**real plots to display** = 20

Displays the next 20 real values of the two output, from the initial location of the cursor *from*. In this case it will be the last 20 real values that will be displayed, the table not containing enough records after the initial location 74 of the cursor *from*.

**0/1 view** = no

Real data and forecasts will be displayed based on their true value.

**to do** = learn & forecast

Neural networks are reset, a step-by-step learning and forecasting process is performed.

## Learning

- ▶ Press the *start* button in the console.
- ▶ Check and confirm the settings.

```
running mode:

database : SCIO-MACHINE-DEMO-
DATABASE.sqlite
table : forex
from : 74
forecast ahead : 3
with past : 4

G : 10
deep : 100
back train set : 2
stretch : false
forecasts to run : 12
real plots : 20
run todo : learn & forecast

view : real range

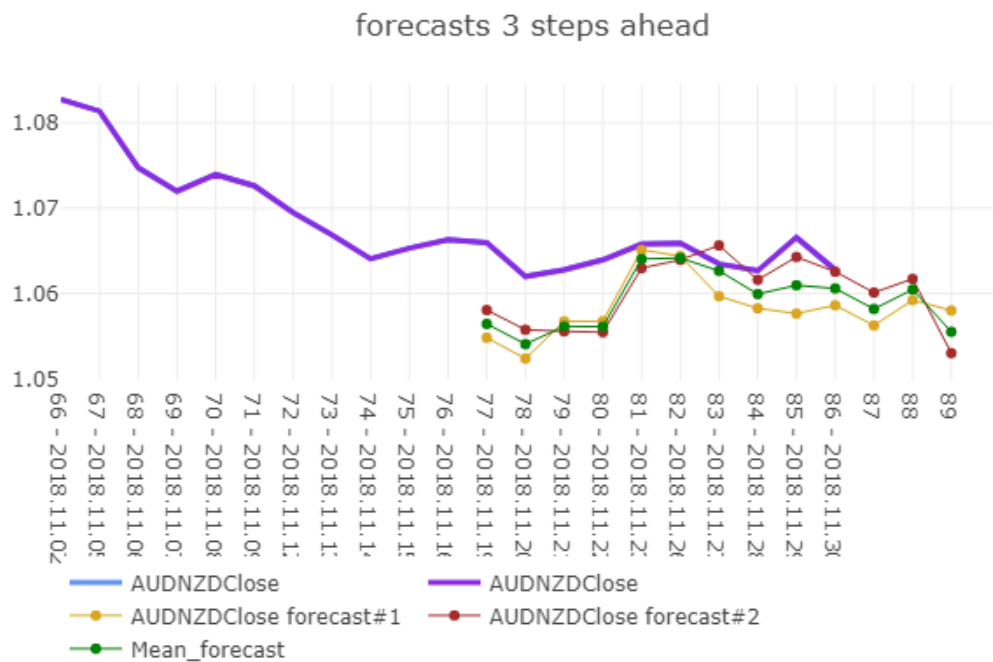
#1 field output : ["AUDNZDClose"]
#1 fields input : 18 fields
#1 type : evolution

#2 field output : ["AUDNZDClose"]
#2 fields input : 18 fields
#2 type : evolution
```

[back to settings](#) [ok](#)

The computing time can be longer than one hour, due to the complexity of the networks and the number of input fields.

The steps of the learning process are displayed in the console. The graph shows the real values of the two output (AUDNZDClose) which will be confused as well as the values of the forecasts computed at each step.



## Report

You can print a report to keep track of forecasts and settings.

► Press *save report*.

Each forecast is calculated three upstream records with the associated neural network as established at the step. Initial results show that the neural network requires several steps before proposing relevant values.

It should be noted that fluctuations between one currency and another do not follow a precise logic and are therefore not very predictable. The consideration of precedence with short past and a limited set of back train tests is relevant in this case.

The two forecasts are calculated with the same parameters and the same input and output fields but their results differ slightly, which gives the user a high and low interpretation range. The green line is the average value.

Given the chronological nature of the data, the forecasts obtained in the *learn & forecast* phase are preferable to those obtained with *forecast only* because each one was produced by the most recent state of the neural network. However a forecasting phase without learning *forecast only* can be made after to know if the last state of the two neural networks produces relevant forecasts upstream or downstream, which may show the existence of some logic in the fluctuation of the AUDNZD pair.

## Forecasts from the latest state of neural networks

Make a forecast without learning.

► Change the settings as follows:

**Forecasts to run:** 20

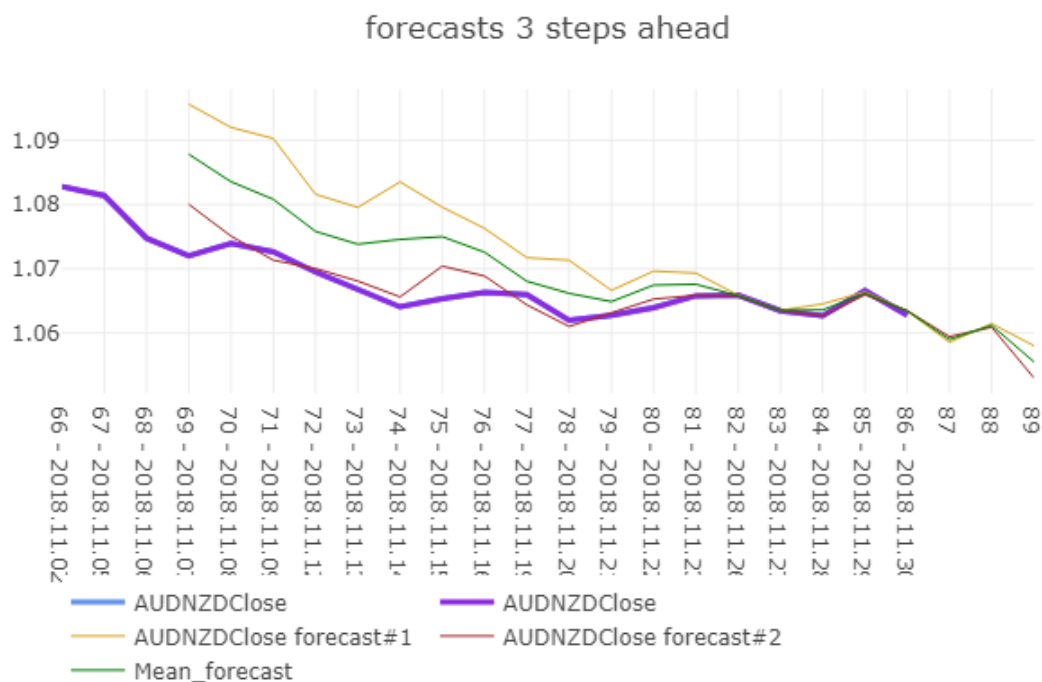
**to do:** forecast only

► Click *start* in the console.

► Check and accept the settings.

A process of calculation of forecasts begins. It uses the acquired state at the end of the *learn & forecast* process of both neural networks.

The result obtained at the end of the process may be different from the graph below, this is due to the very character of the genetic selection algorithm. However it must have strong similarities.



It is noted that the forecasts obtained by network #2 are more relevant and suggest a certain logic.

The calculated forecasts show a downward trend at three days (quotation).

This example has been produced with the aim of showing how Scio-machine.com works as part of the forward forecast. Any forecast carries a notion of uncertainty, even more so in the market of forex deemed to withstand the calculations of prediction. It is important to cross-reference forecasts with other assessments before making responsible decisions.

Mechanized prediction must be used with a benevolent ethic towards the human being and his interests.

## Features & Use

### **How Scio-machine.com works**

Scio-machine.com is a forecasting software. Supervised machine learning techniques with neural networks and genetic selection algorithms are implemented.

It is easy to use and lets the user focus on objectives and data.

The data should be in Sqlite format, very easy to produce from a spreadsheet file or CSV.

Scio-machine.com can be used to produce forward forecasts from time and multi-factor (attributes) data series .

Scio-machine.com can also be used from multi-factor (attributes) data collections such as a questionnaire to select or determine a profile or to know the impact of certain factors (attributes).

Scio-machine.com simultaneously produces two independent forecasts based on data from the same table, and produces two neural networks.

Learning and forecasting are done in successive steps.

### **Preparation of the data**

Considering your data collected in a CSV file.

#### Check these points:

- + If it is a time series records the oldest must be the first (line 1) and the most recent the last.
- + Only numerical value fields (columns) will be used.
- + The decimal separator must be the point.
- + Name the fields with alphanumeric characters without accents, special signs or spaces.
- + There must be no zero value or empty cell in the numeric fields.

#### Conversion of labels to numerical values

It may be that some of the data you want to use happens to be labels (text), so you will have to transform these labels into numerical values. For example “Blue”, “Red”... in 1, 2...

DataPreparator (<http://www.datapreparator.com>) is a proprietary software that is available free of charge and will help you greatly in this task, and more generally in the preparation of your data.

## Conversion of CSV file to Sqlite file

You can convert your CSV file with a Sqlite database manager such as Open source DB Browser for sqlite (<https://sqbrowsliteer.org/>).

- + Create a new database and import the CSV file as a table.
- + Modify the table by changing the type of numeric fields to *REAL*.
- + Make sure that the table name contains only alphanumeric characters, without spaces or accents.
- + Save your changes, your Sqlite file is ready.

### **More info/advice**

- + The x-axis (horizontal axis) of the graph corresponds to the order numbers of the records. The origin is 0 (the first record).
- + The record numbers displayed on the x-axis can be enriched with other information from a field called "scio\_X" and containing descriptions.
- + The output field must not be selected in the input field if it is not a forward forecast, ***forecast ahead = 0***.
- + Classification and Evolution modes refer to two different ways of calculating a forecast, and can be compared by metaphor to predict the Altitude of the next summit or the difference between the current and the next summit. The ***Evolution type*** is only available for forward forecast, ***ahead > 0***.
- + The cursor ***from*** can be automatically moved at startup depending on the conditions. For example if ***from*** is placed on the last record.
- + Profilers are not available for forecasts calculated with input fields spread over more than one record, ***with past > 1***, or with the ***type = Evolution***.
- + When converting labels (text in a column) into numerical values, try to respect the semantic order and if possible the semantic deviation. The best still being to use measurable characteristics, for example colors can be replaced by their respective light frequency.
- + It is not advisable to separate the data, for example to separate the data for learning from the data for forecasting. The learning setting allows you to select the part(s) of a table for this task. If you have to separate the data, make sure that the set used for learning contains representative minimum and maximum values for each field.
- + Scio-machine.com is not a statistical software but a forecasting software. In other words, statistics make it possible to compare and the forecast must make it possible to understand. Therefore, the data must reflect circumstances or characteristics.

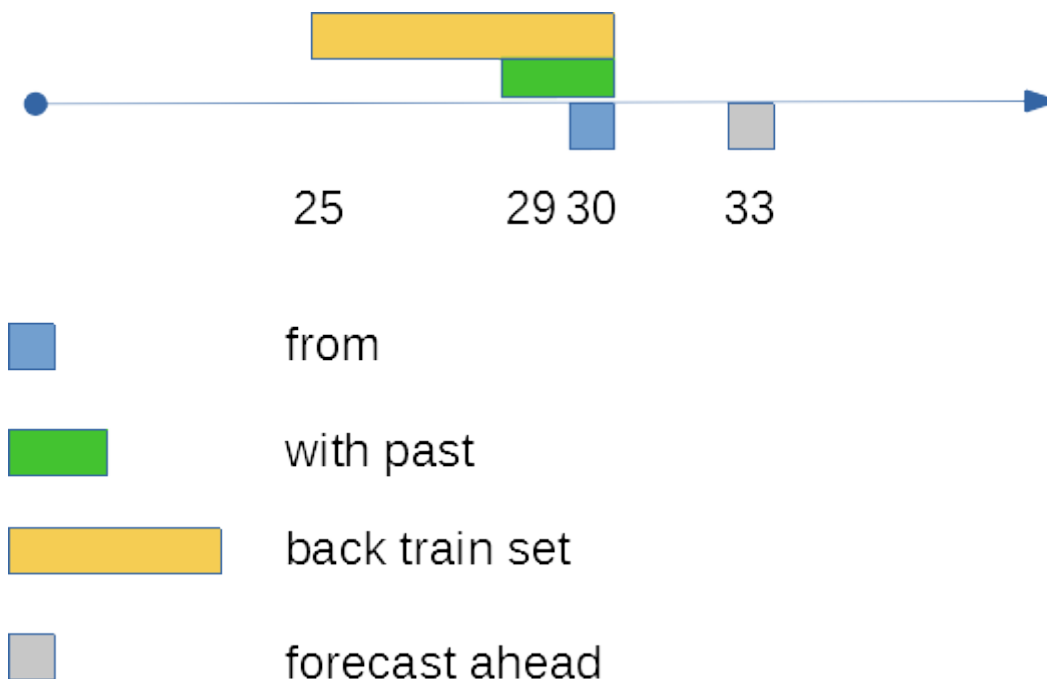
- + Watch out for false testimonies! Make sure that the data in a field reflects the same reality. For example, the field examScore should not have a 0 if the candidate is absent from the examination. This does not reflect his level and may therefore mislead a learning error.
- + Learning strategies can be different depending on the situation. In general, make sure you take into account cycles and enough different situations.

To help you the following diagram shows the implication of certain parameters during the *learn & forecast* and *forecast only* steps.

In the mode **to do** = *learn & forecast* the input fields in the records covered by **back train set** (25 to 30) are used to learn and calculate the output field forecast from the **forecast ahead** 33 record. At the next step, all positions will be incremented by 1.

In the mode **to do** = *forecast only* the input fields in the records covered by with past (29-30) are read to calculate a forecast of the output field in the **forecast ahead** 33 record. At the next step, all positions will be incremented by 1.

**from** = 30  
**with past** = 2  
**back train set** = 2  
**forecast ahead** = 3



## Security & privacy

Scio-machine.com works locally on your computer. No transfer of your data takes place. You need an internet connection to download the settings page. Once loaded it is possible to continue working offline.

No personal data is stored on the server. There is no password for using the software, even in corporate mode. Scio-machine.com does not use cookies and does not store anything on client side devices.

## License

Scio-machine.com can be used in community or corporate mode.

The “community” mode is free and subject to a restriction on the number of input fields. Access to the online software is direct from <http://machinie1.scio-machine.com> with a recent Chrome or Firefox browser.

The "corporate" mode is paying and unrestricted. Access to the online software is from a launcher to download from <http://scio-machine.com> . You need a recent Chrome or Firefox browser. The "corporate" license and its sub-licenses are each one attached to a customer device.

## Parameters

| Text of the parameter | Function  |
|-----------------------|---|
| <i>Choose a file</i>  | Allows to select the database to use. This database will never be modified or transferred to a server. The database must be in SQLITE format, it can contain several tables.  |
| <i>table</i>          | Select the database table containing the records to be used.  |
| <i>inputs</i>         | List of input fields used for learning and forecasting. Only fields containing numerical values are available.  |
| <i>output</i>         | Output field name for the forecast. Only fields containing numerical values are available.  |
| <i>validate</i>       | Click to validate the choices.  |
| <i>type</i>           | 2 methods of forecasting calculation are proposed.<br><br>By analogy, <i>Classification</i> is based on altitude and <i>Evolution</i> is based on altitude difference.<br><br><i>Evolution</i> cannot be used for a zero anticipation, <i>forecast ahead</i> = 0. |
| <i>from</i>           | Refers to the record from which the forecasts will begin. <i>from</i> is a cursor.<br>The starting position can be adjusted automatically if the settings require it.   |
| <i>with past</i>      | Anteriority in number of records. These records will provide the  |



|                              |  |
|------------------------------|--|
|                              | <p>input values for the forecast calculation, at each step.<br/> 1 is equivalent to the record <b>from</b>.<br/> 4 equals the record pointed by <b>from</b> and the 3 preceding it.</p>  |
| <b>forecast ahead</b>        | <p>Magnitude of the future, in records, for the calculation of forward forecasts.<br/> 3 is equivalent to calculating forecasts at 3 records further than the one pointed by <b>from</b>.<br/> 0 is equivalent to calculating the forecast for the record pointed by <b>from</b>.</p>  |
| <b>G</b>                     | <p>Number of genetic evolutions produced at each step.<br/> Calculation time increases with <b>G</b>.</p>  |
| <b>back train set</b>        | <p>Set of tests to perform at each step of learning. This set uses the records immediately preceding the record pointed by <b>from</b>.<br/> 1 is the minimum.</p>   |
| <b>forecasts to run</b>      | <p>Number of steps. <b>from</b> is incremented by a position at each step.<br/> 1 is the minimum and produces 2 forecasts.</p>   |
| <b>stretch</b>               | <p>If checked, add the previous test sets to the new ones.</p>   |
| <b>deep</b>                  | <p>Increases the complexity of neural networks. Allows greater predictability.<br/> Default: 1. Maxi: 1000.<br/> Consumes more memory and computing time, and can cause browser interruption.</p>  |
| <b>real plots to display</b> | <p>Number of real values of the output fields to be displayed.</p>   |
| <b>0/1 view</b>              | <p>Determines the display scale to be used on the graph.</p> <p>“<b>No</b>” the data will be displayed at their real value.</p> <p>“<b>yes</b>” the data will be displayed normalized between 0 and 1.<br/> This option allows to visually compare data with each other that have a very large difference scale in value.</p>  |
| <b>to do</b>                 | <p>Action to be taken.</p> <p>“<b>learn &amp; forecast</b>” resets neural networks and launches the learning and forecasting phase simultaneously. Learning takes place in steps. At each completed learning step a forecast is calculated and displayed. If <b>forecast ahead</b> = 0 then the forecast should be recalculated at the end of the process with <b>forecast only</b>.</p> <p>“<b>forecast only</b>” only performs prediction calculations based on the state of the neural networks acquired at the end of the learning phase of <b>learn &amp; forecast</b> or <b>learn more</b>. These forecasts can be calculated on another part of the table, or on another table. Changes to certain parameters can lead to inconsistencies in results.</p> <p>“<b>display only real</b>” displays only the real values of the two output fields. Useful for making comparisons 2 by 2. Note that a selection of input fields must be made but it will not be used.</p> |

|                          |   |
|--------------------------|---|
|                          | “ <b>learn more</b> ” carries out an additional phase of <i>learning &amp; forecast</i> without resetting neural networks. Useful to produce learning on multiple portions of a table containing a large number of records. |
| <b>save report</b>       | Displays a report to print. It integrates the last graph obtained as well as the last settings of the parameters.   |
| <b>demo &amp; help</b>   | Allows the download of the demo database and guide books.   |
| <b>PROFILE #1 and #2</b> | Allows to produce a forecast based on a custom adjustment of input fields.<br><br><b>PROFILER</b> is only available for forecasts configured with <b>with past</b> = 1 and <b>type</b> = classification.                    |

Any forecast carries a notion of uncertainty. It is important to cross-reference forecasts with other assessments before making responsible decisions.  
Mechanized prediction must be used with a benevolent ethic towards the human being and his interests.