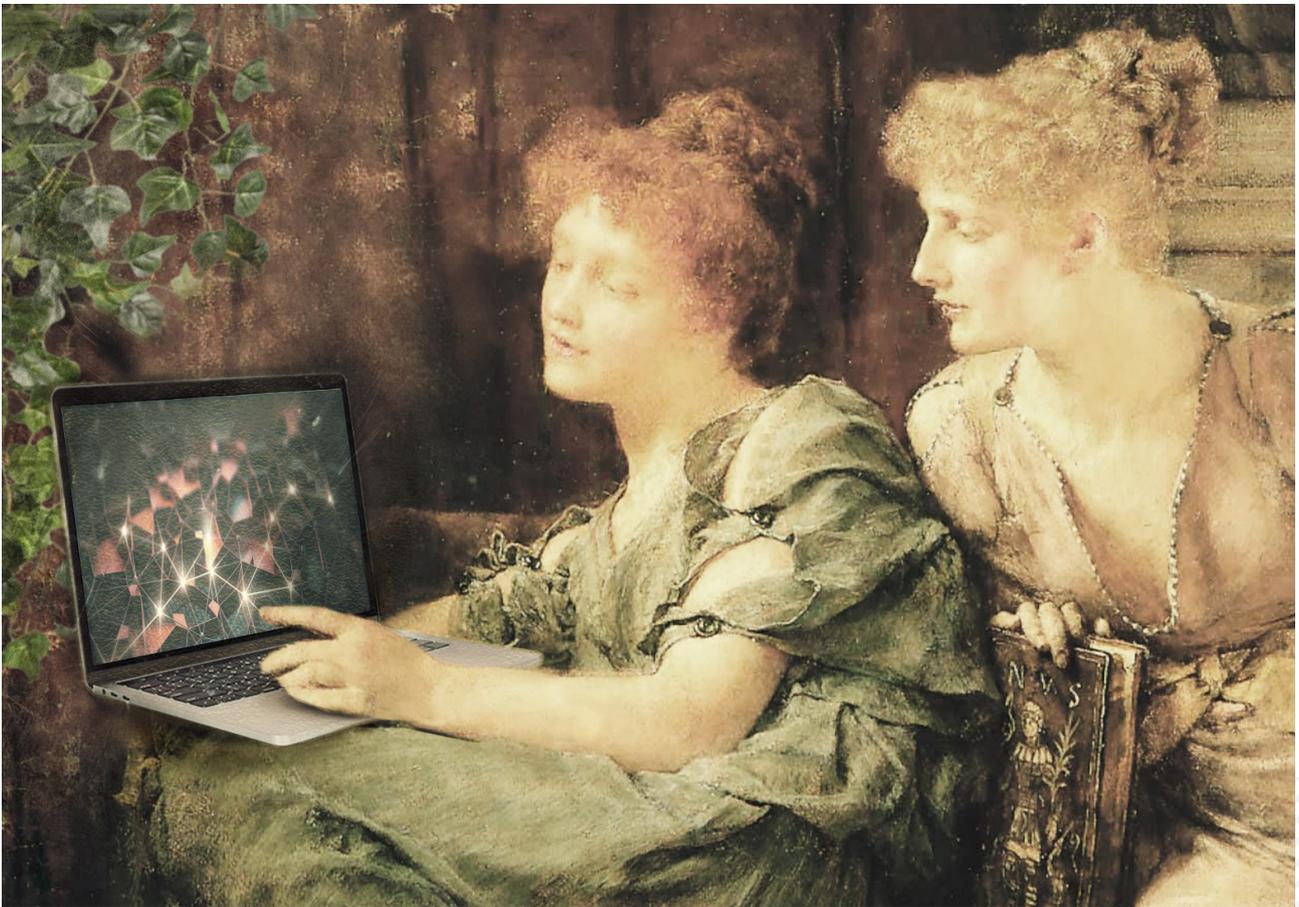


# Neuton Quick Start Guide



# CONTENT

|                                       |           |
|---------------------------------------|-----------|
| <b>Log in to Neuton</b> .....         | <b>3</b>  |
| <b>Select data for training</b> ..... | <b>4</b>  |
| <b>Train your model</b> .....         | <b>8</b>  |
| <b>Make predictions</b> .....         | <b>10</b> |

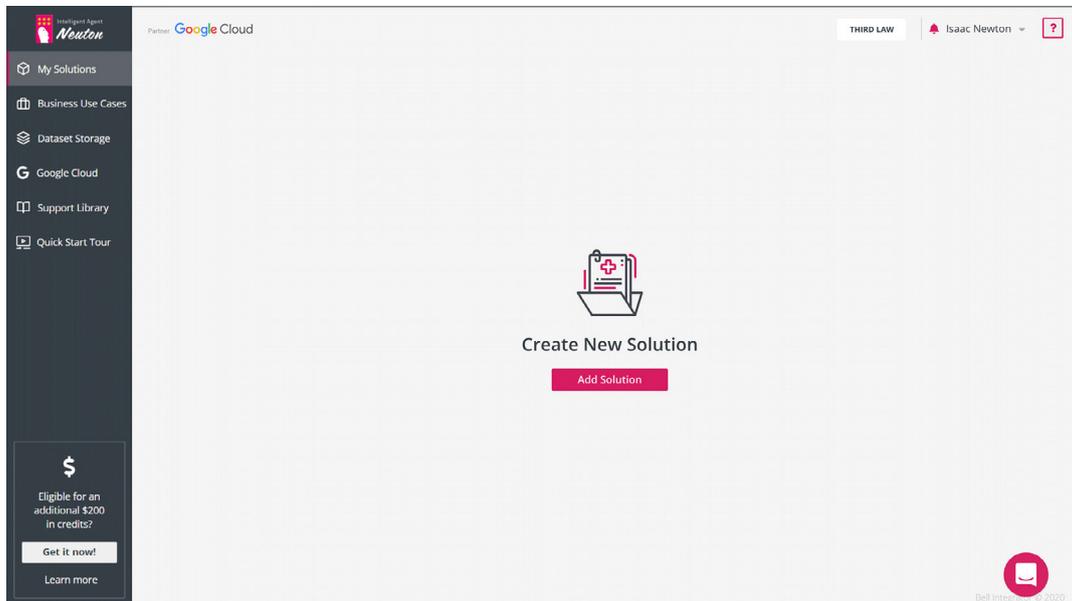
---

## LOG IN TO NEUTON

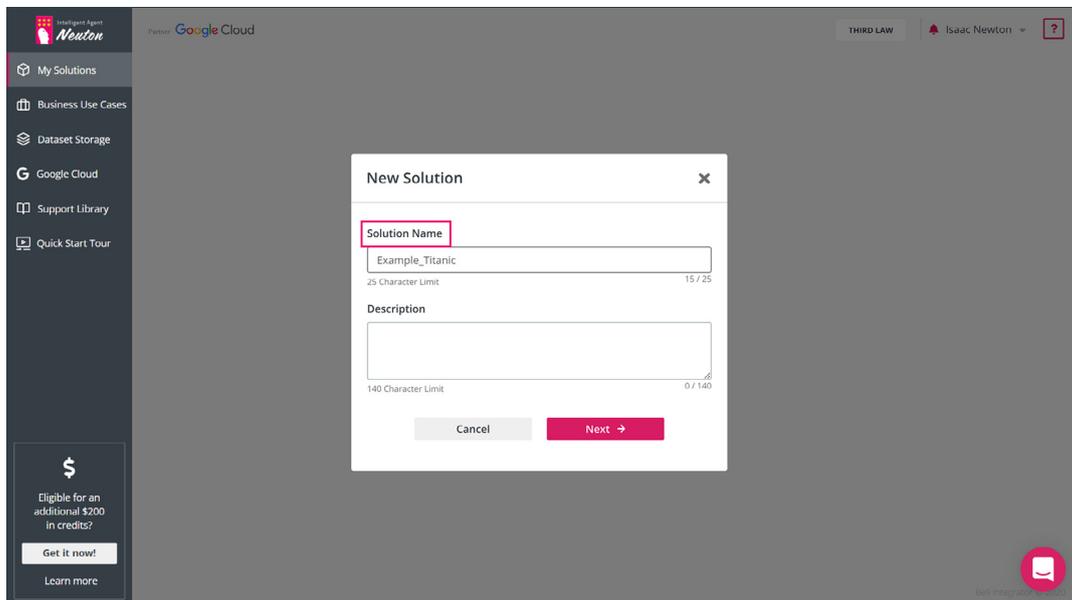
Go to <https://neuton.ai> in your web browser to access Neuton. After page refresh click «**Sign in**» to start using Neuton. You will be redirected to the «**Sign in with Google**» web page, after that please provide the **email** address and **password** used during Neuton registration.

# SELECT DATA FOR TRAINING

Create your first solution using the «Add Solution» button.

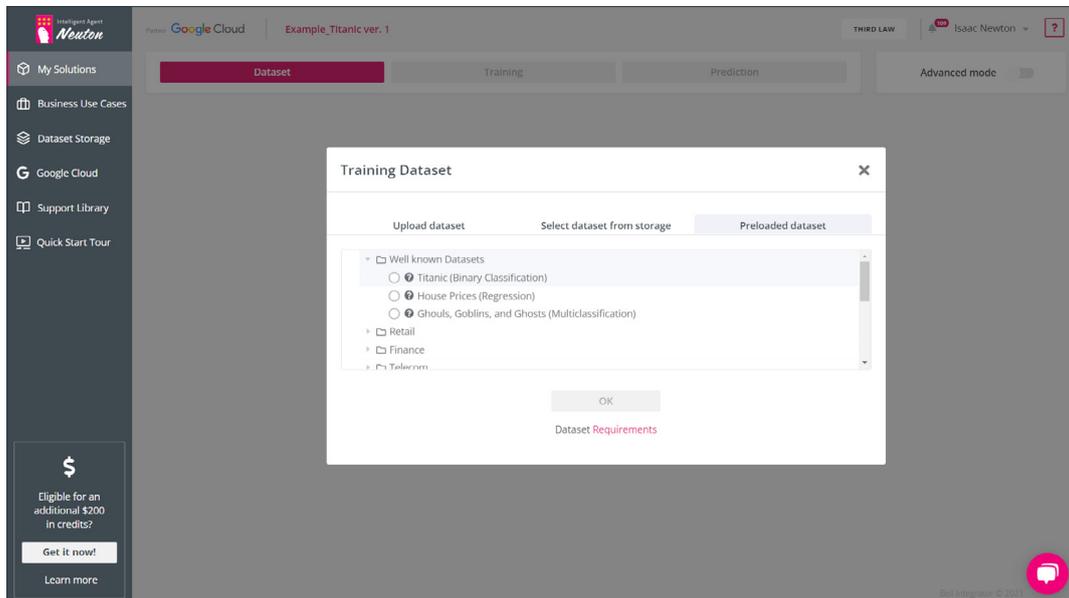


Specify the «Solution Name» in the pop-up window.



Press the «Next» button to select the preloaded dataset or upload a new one.

On the next screen press the **«Select Dataset»** button. Upload training dataset from the Kaggle Titanic ML competition or navigate to the **«Preloaded dataset»** tab and choose **Titanic**:



Press the **«OK»** button to confirm selected dataset.

## NOTE:

### List of requirements for training datasets in Neutron

1. Dataset must be a CSV file using UTF-8 or ISO-8859-1 encoding.
2. Dataset must have a minimum of 2 columns or rows.
3. Dataset must not have any empty columns.
4. Pay attention to duplicate rows: Duplicate rows will be removed from the training set during the preprocessing stage, which may cause an inconsistency with the dataset minimum size and sample number requirements.
5. The first row in the dataset must contain the column names, and a comma, semicolon, pipe, caret or tab must be used as a separator. CRLF or LF should be used as the end of a line character. The separator and end of line character should be consistent inside the dataset.
6. All column names must be unique and must contain only letters (a-z, A-Z), numbers (0-9), hyphens (-) or underscores (\_).
7. All empty values must be excluded from the target variable.
8. For regression task types the target variable must have only numeric values.
9. For classification task types at least ten samples must be provided for each class of target variable.

10. Currently Neuton supports only the EN-US locale for numbers and timestamps, so:

- You must use a dot as a decimal separator, and delete spaces and commas typically used to separate every third digit in your numeric fields.
  - Example – "20,000.00" should be replaced with "20000.00"
- If any numeric column is represented as a combination of a number and its corresponding unit, then only the number should be placed in the column.
  - Example – "\$20,000.00" should be replaced with "20000.00"
- You can use any of the following date/timestamp formats:
  - "DD.MM.YYYY hh:mm:ss"
  - "MM.DD.YYYY hh:mm:ss"
  - "YYYY.MM.DD hh:mm:ss"
  - "YYYY.DD.MM hh:mm:ss"
  - "DD.MM hh:mm:ss"
  - "MM.DD hh:mm:ss"
  - "MM.YYYY hh:mm:ss"
  - "YYYY.MM hh:mm:ss"

You can use "/" ";" or "." as date components separator. Space should be used as date and time separator and ":" should be used as time components separator.

- The timestamp format should be consistent inside any given column.
- In test datasets the same timestamp format should be used in a manner consistent with the training dataset.

11. Text fields should start and end with quotes, in order to avoid confusing symbols in the text with column separators.

12. End of line symbols must be excluded from the fields values.

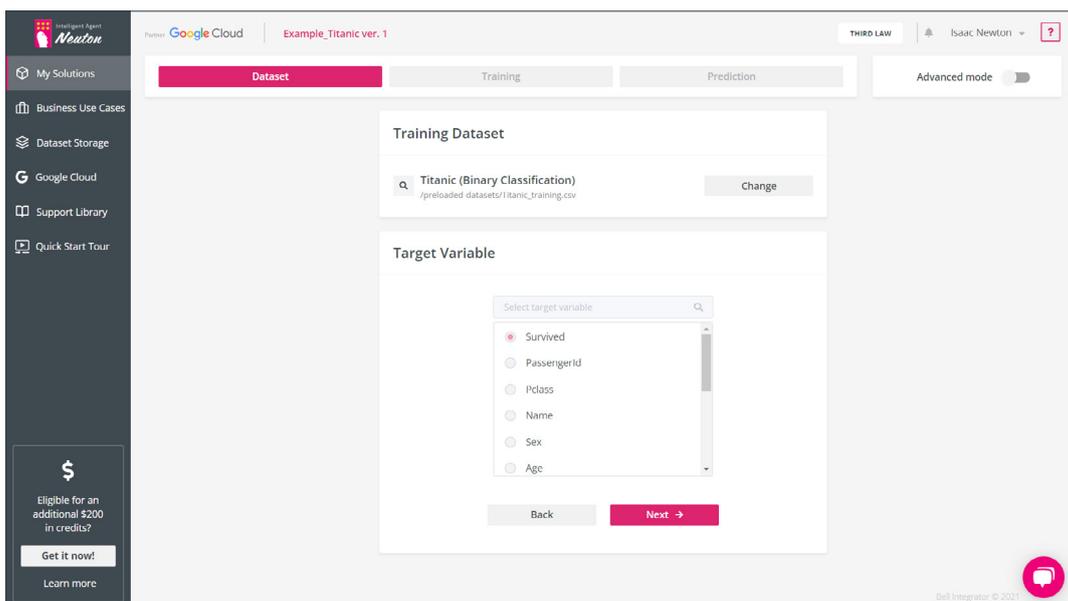
13. Neuton supports NLP, but can currently only process textual fields in English (US). If you used a language different from English for categorical fields in your dataset, you must transliterate them into English (US).

#### **List of requirements for test datasets (or new data)**

1. Dataset must be a CSV file using UTF-8 or ISO-8859-1 encoding.
2. The first row in the dataset must contain the column names, and a comma, semicolon, pipe, caret or tab must be used as a separator. CRLF or LF should be used as the end of a line character.
3. The test dataset must have the same file structure and data types as the training dataset.

4. The column names must be the same and the columns must have the same order as in the training dataset.
5. The target variable column and its name must be excluded from the test dataset.
6. For the Gravity Subscription Plan the size of your test set should not exceed 100 rows.
7. In test datasets the same timestamp format should be used in a manner consistent with the training dataset.
8. Text fields should start and end with quotes, in order to avoid confusing symbols in the text with column separators.
9. End of line symbols must be excluded from the fields values.

### Select **Target Variable**.



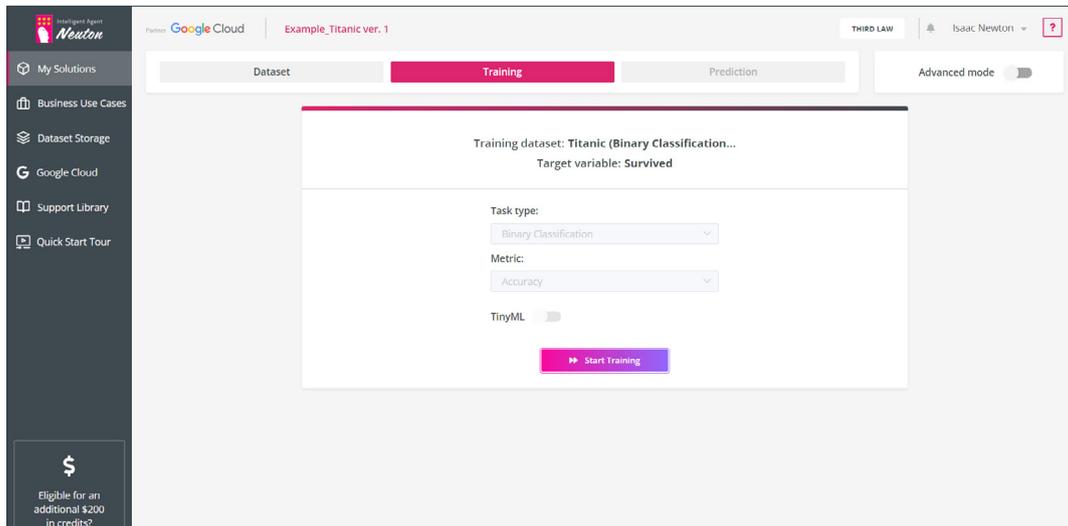
Press the **«Next»** button to go to next step.

### **NOTE:**

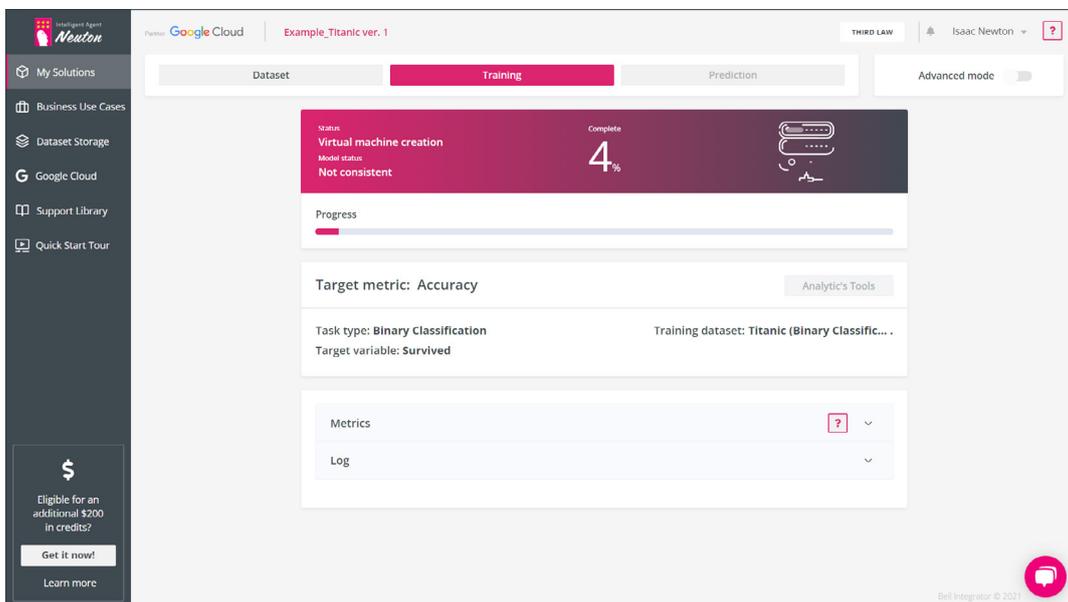
For the preloaded datasets all settings are preconfigured.

# TRAIN YOUR MODEL

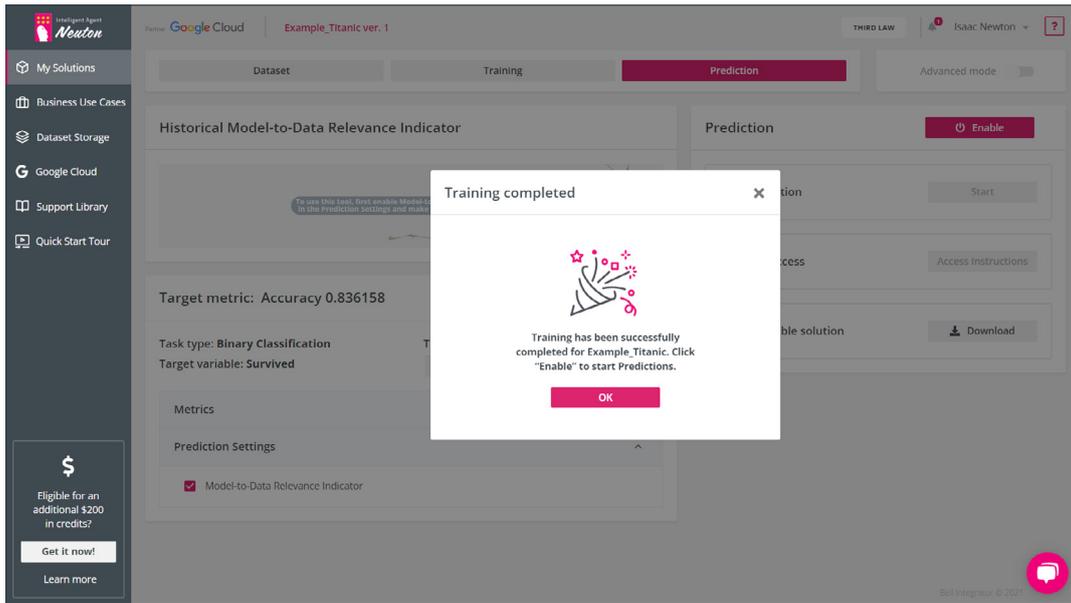
On this step **Neuton** will automatically offer you **Task type** and target **Metric** but you can change it according to your preferences. To continue press **«Start Training»**.



You can track the percentage of training completion on the progress bar at the top of the training tab.



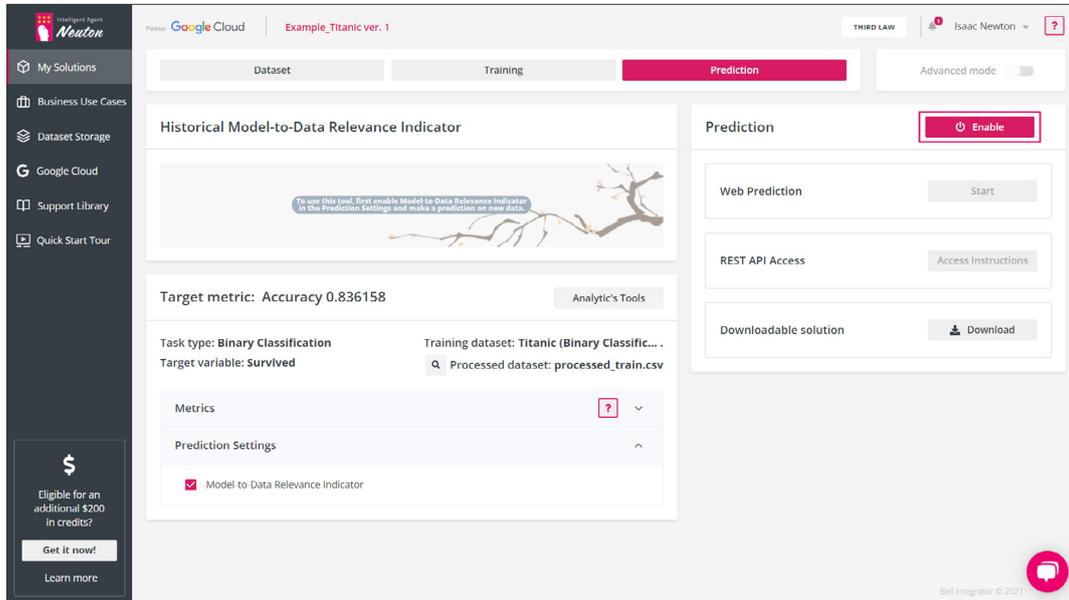
As a result of training, the model is created.



Press the «OK» button to go to Prediction stage.

# MAKE PREDICTIONS

When the training is completed click the «Enable» button to switch on prediction.



## NOTE:

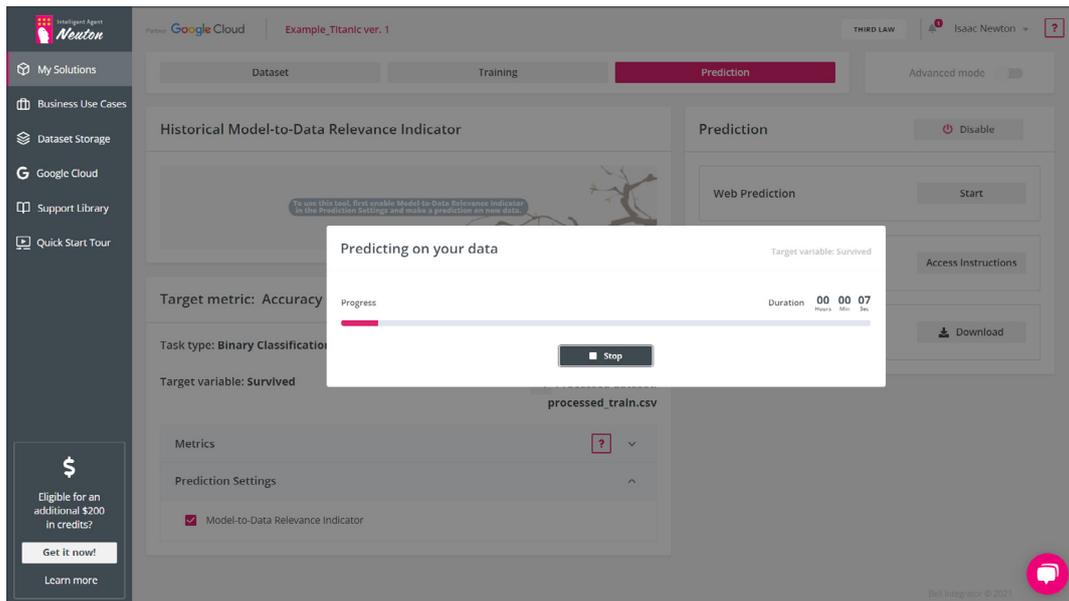
That additional costs may be charged for a new virtual machine for predictions.

When the Virtual machine is created, the «Start» button will be activated. Press this button to start **Web Prediction**. In the pop-up window click «Start» (for the preloaded datasets the test dataset is preselected). If you haven't used the preloaded dataset you can upload your test dataset (or new data) without a target variable to make the prediction (this process is similar to training dataset selection). You can download a test dataset from the Kaggle Titanic ML competition.

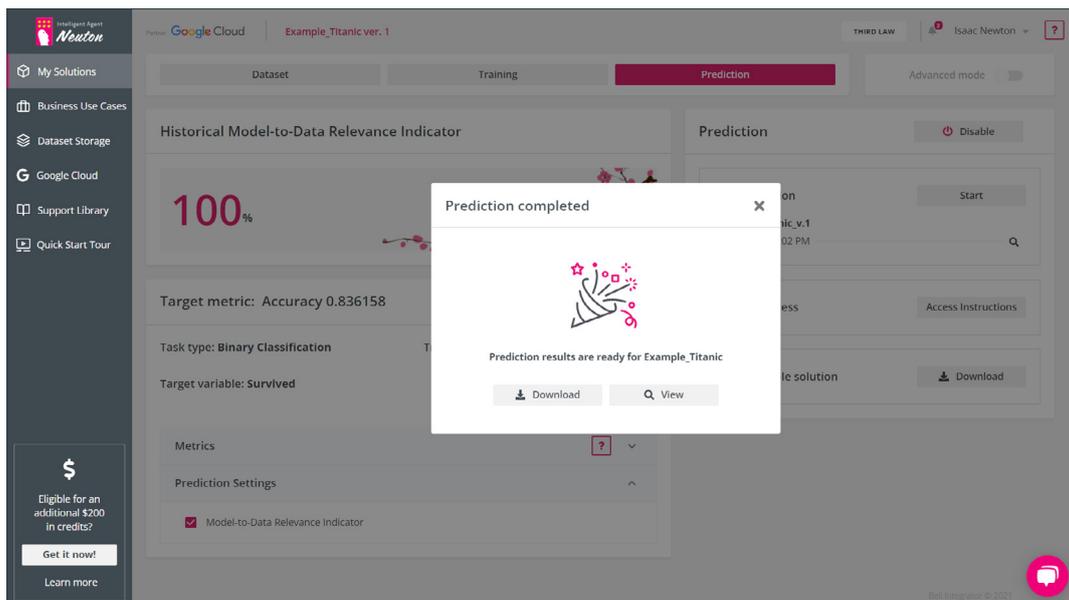
## NOTE:

The test dataset structure and variable types should be the same as in the training dataset, except that the target column will be excluded.

Wait for predicting completion. It can take a few minutes.



Congratulations! Your predictions are completed!



Download prediction results for further insights!

For more information about Data Science and improving prediction results on your dataset please refer to [Neutron User Guide](#).