FARRAGO

SUPERCHARGE AI PRODUCTIVITY

Machine Learning
Results Report

21.05.2019

Farrago Machine Learning Results

Farrago Machine Learning lets you easily make use of complex machine learning algorithms to predict data in your business. Based on our analysis of your data, it is transformed and formatted into machine readable formats. We then use machine learning algorithms to learn and predict new things about your data and find insights you may not have foreseen.

Farrago ML Run Overview

For this run, you chose to perform *Match outcome prediction* over your data.

Your data was analysed based on the *linear_learner_classification* algorithm as well as your selected column of interest (*FTR (target)*), and some data preprocessing transforms were recommended to improve the accuracy of the final results from this run.

After the preprocessing was completed, your data was used to train a *linear_learner_classification* model, which was then used to predict values that were missing from your selected column.

This report presents the results from this machine learning run, including the accuracy of the learned model based on a validation set selected from your data, and a summary of the preprocessing transforms that were recommended and chosen.

Sport_Match_Outcome_Prediction_Classification.csv

|	
Date processed	21.05.2019
File size	42 kb
Prediction type	Match outcome prediction
Algorithm used	linear_learner_classification
Values analysed	380

Results

When it was run over your data, the *linear_learner_classification* model achieved a validation accuracy of

89%

In order to obtain this accuracy, Farrago ML selected a tenth of your data for validation purposes and removed it from the model training process. It then used the model to predict values in the *FTR (target)* column and compared these results against the original values in the validation data.

Here is an overview of the results from the model validation process.
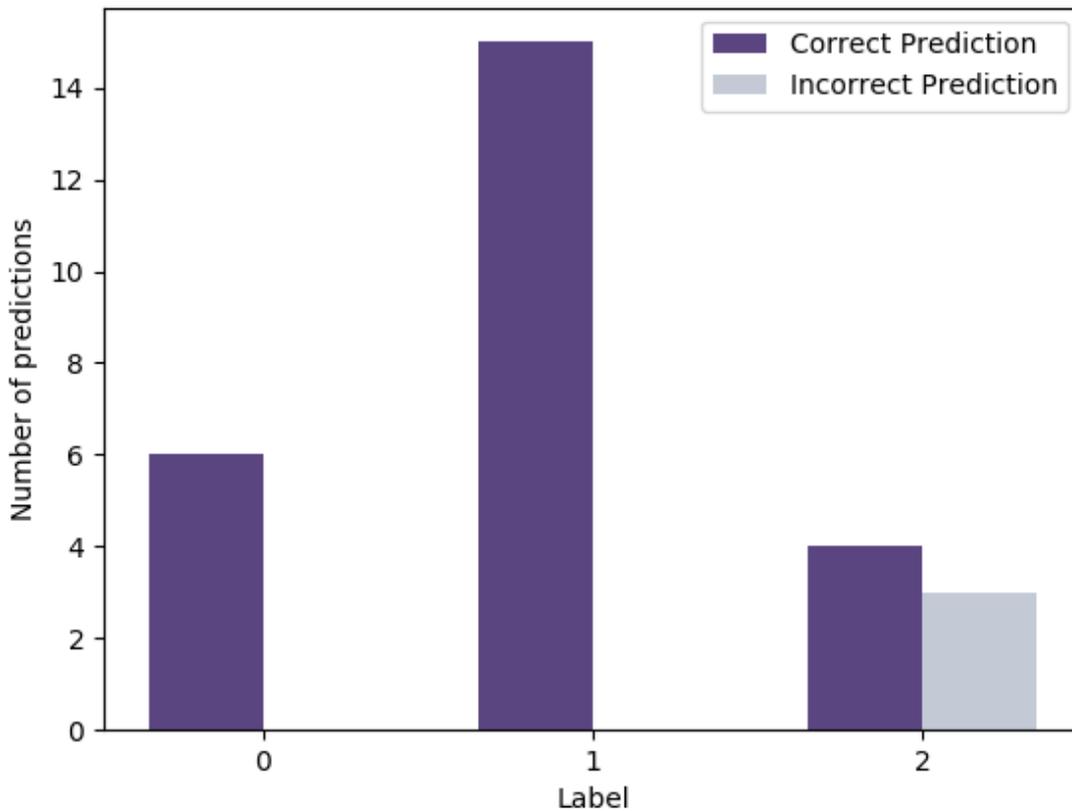
This graph shows the number of values that were predicted to be each value. The purple bars represent the number of values that were correctly predicted as each value, and the grey bars represent the values that were predicted as that value, but where the actual value was something else.

The following table shows the first few rows of the data that these results are based on.

Original Validation Data	Predicted Value
FTR (target)	**FTR (target)**
A	A
A	A
A	A
A	A
A	A
A	A
H	H
H	H
H	H
H	H

Your output results can be downloaded from https://ml.farrago.ai/view-results/0f86f4567a407b5cfc4e/2019-05-22T10-07-28, for up to 2 hours after first accessing the page. After that, your data will be removed from our systems to avoid security concerns.

Data Preprocessing

Farrago ML has a number of data transforms that it applies in order to make your data more usable and useful for machine learning. Some of the transforms are required in order for Farrago ML to complete the run, while others can be enabled or disabled at your discretion to improve the accuracy of your results.

Selected Transforms

The following table contains transforms that were recommended based on the Farrago ML data analysis process and that you chose to use. They address issues found within your data for the selection of *Match outcome prediction* over *FTR (target)*.

Most of the time, performing these steps will help in improving the accuracy of the final results from machine learning.

Transform	Issue the Transform Addresses	Effects of the Transform
Column Scaling	Columns "FTHG", "HTHG", "HTAG", "HS", "AS", "HST", "AST", "HF", "AF", "HC", "AC", "HY", "AY", "AR", "var_1", "var_2", "var_3" have unusual scaling.	This can affect the performance of many algorithms, it is recommended to normalise them.

Required Transforms

The following table shows transforms that address a number of fundamental issues that Farrago ML found within your data. These issues relate to how machine learning data must be formatted in order to be readable by SageMaker.

Transform	Issue the Transform Addresses	Effects of the Transform
Missing Values	There are missing values spread throughout the data.	We recommend imputing them with column medians (for numeric data) or a special tag (for categorical data).
Numeric Features	Some columns are composed of text data	Sagemaker requires text data be made numeric , columns ['Date', 'HomeTeam', 'AwayTeam', 'HTR', 'Referee', 'Time of match'] will be transformed into nu...
Numeric Labels	Class labels in the target column (FTR (target)) is text data.	Sagemaker requires numeric representations of the class label, so they will be transformed into values from 0 to 2

Farrago ML uses SageMaker to run its machine learning functionality, and so some restrictions are placed on the format of

certain data types. These transforms address those restrictions and fit your data to SageMaker's input requirements.

However, as many of the transforms address fundamental issues in how machine learning algorithms understand data, it is likely that a Data Scientist would recommend some or all of these transforms to be applied to your data anyway.

Farrago Machine Learning Process Logs

```
[2019-05-22 10:07:28]  Copying demo file to current run's folder

[2019-05-22 10:07:29]  Copied demo file successfully to current run folder

[2019-05-22 10:07:29]  Performing basic data pre-formatting

[2019-05-21 22:07:32]  Performing generic analysis on user's data

[2019-05-21 22:07:32]  Data not empty and not too large

[2019-05-21 22:07:32]  Analysing what functionality each column can be used for

[2019-05-21 22:07:32]  Finished generic dataset analysis

[2019-05-22 10:07:33]  Preparing use case list

[2019-05-22 10:07:33]  Filtering use cases by demo file

[2019-05-22 10:07:33]  Presenting dataset view for use case/algorithm selection

[2019-05-22 10:07:43]  User selected use case 'Match outcome prediction' - use
                       linear_learner_classification algorithm

[2019-05-22 10:07:43]  User selected column '5' as target

[2019-05-22 10:07:43]  Setting up run metadata

[2019-05-22 10:07:43]  Finished setting up run metadata

[2019-05-22 10:07:43]  Performing in-depth analysis of user's data

[2019-05-21 22:07:47]  Starting analysis of client's data

[2019-05-21 22:07:47]  Data size: 380 rows

[2019-05-21 22:07:47]  Building recommendations for data transforms based on the selected
                       algorithm and data.

[2019-05-21 22:07:47]  Finished building recommendation list.

[2019-05-21 22:07:47]  Determining required data transforms for the selected algorithm and data.

[2019-05-21 22:07:47]  Finished determining required data transforms.

[2019-05-21 22:07:47]  Formatting recommendations to display to user

[2019-05-21 22:07:47]  Recommendations formatted

[2019-05-22 10:07:47]  Finished performing in-depth analysis
```

[2019-05-22 10:07:47]	Preparing recommendations for user
[2019-05-22 10:07:47]	Displaying recommendations to user
[2019-05-22 10:07:53]	Setting up analysis report PDF
[2019-05-22 10:07:53]	Analysis report variables set up
[2019-05-22 10:07:53]	Building analysis report PDF
[2019-05-21 22:07:54]	Analysis report PDF built
[2019-05-22 10:14:09]	User selected the following recommended transform: column scaling
[2019-05-22 10:14:09]	Performing data transformations based on user's selection
[2019-05-21 22:14:13]	Preparing to perform data transforms over user's data
[2019-05-21 22:14:13]	Building transform pipeline
[2019-05-21 22:14:13]	Transform pipeline built
[2019-05-21 22:14:13]	Running through pipeline to perform data transforms
[2019-05-21 22:14:14]	Finished transforming data
[2019-05-21 22:14:14]	Splitting transformed data into train, validation and output sets
[2019-05-21 22:14:14]	No missing values found in target column, using 25% of data as example output set
[2019-05-21 22:14:14]	Splitting a representative sample of the training data for validation of the model's accuracy
[2019-05-21 22:14:14]	Finished splitting the transformed data
[2019-05-21 22:14:14]	Finished transforming data for SageMaker
[2019-05-22 10:14:15]	Displaying transform results to user
[2019-05-22 10:14:18]	Starting Machine Learning processes in SageMaker
[2019-05-21 22:14:21]	Starting AWS SageMaker training job
[2019-05-21 22:18:27]	Building SageMaker model
[2019-05-21 22:18:27]	SageMaker model built
[2019-05-21 22:18:28]	Preparing SageMaker job for validation
[2019-05-21 22:18:28]	Running validation job to test model accuracy
[2019-05-21 22:22:33]	Validation job completed

```
[2019-05-21 22:22:33]   Model accuracy: 89%

[2019-05-21 22:22:33]   Uploading validation statistics to S3

[2019-05-21 22:22:37]   Starting SageMaker inference job.

[2019-05-21 22:26:43]   SageMaker output detected - finalising results

[2019-05-21 22:26:43]   Making raw results readable and merging with original data

[2019-05-21 22:26:43]   Building data visualisation of the validation output

[2019-05-21 22:26:44]   Data visualisation built

[2019-05-21 22:26:44]   Uploading results to S3

[2019-05-21 22:26:44]   Results url: https://ml.farrago.ai/view-
                        results/0f86f4567a407b5cfc4e/2019-05-22T10-07-28

[2019-05-21 22:26:44]   Building results report PDF

[2019-05-21 22:26:44]   Results report PDF built
```