FARRAGO

SUPERCHARGE AI PRODUCTIVITY

Machine Learning
Analysis Report

22.05.2019

Farrago Machine Learning Analysis

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 forseen.

Analysis Overview

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 the following report presents the findings on what work needs to be completed in order to make your data ready for machine learning.

Sport_Match_Outcome_Prediction_Classification.csv

Date processed	22.05.2019
File size	42 kb
Prediction type	Match outcome prediction
Algorithm used	linear_learner_classification
Values analysed	380

ESTIMATED TIME SAVED 3.0 days

Recommendations

The following table shows a number of issues that Farrago ML found within your data for the selection of **_Match outcome prediction_** over **_FTR (target)_**.
The table also lists the ways in which Farrago ML can address these issues for you.

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

Issue	Description	Recommendation	Task time
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.	1.0 days

We estimate that it would take a Data Scientist a total of 1.0 days to perform these tasks.

Required Transforms

The following table shows a number of 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.

Issue	Description	Recommendation	Task time
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).	1.0 days
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...	0.5 days
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	0.5 days

We estimate that it would take a Data Scientist a total of 2.0 days to perform these tasks.

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.

Unneeded Transforms

Other than the issues listed before, Farrago ML also checks for a number of other issues within your data.

The issues in the table below are areas where your data passed all checks and no further work needs to be done.

There are, however, more issues that your data was not tested for, as they relate to other configurations of use cases, algorithms and/or target columns.

Issue	Description	Recommendation	Task time
Class Label Outliers	No changes need to be made.	No action required.	-
Target To Front	No transformation required	No action required.	-
Float Features	No transformation neccisary.	No action required.	-

No time required at all, time for lunch!

Conclusion

After analysing your data, Farrago ML has found that performing *Match outcome prediction* over the *FTR (target)* column of your data would take a data scientist around *3.0 days*.

Don't worry though - Farrago ML can perform the same tasks in under two minutes, in a few simple steps.

1. Send your data through Farrago ML now, and it will automatically resolve all of the flagged issues.
2. Then quickly and painlessly get your data into the SageMaker machine learning models and get started with using predictive analytics to better your business.

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