SparkCognition™ Darwin™ Release Notes

v 1.5 - 10.15.2018

Darwin Release Notes v 1.5

Darwin release 1.5 has incorporated customer feedback to provide improvements in the speed and accuracy of model building. The following changes are completed and rolled into the Darwin release 1.5 for immediate use:

New Features in 1.5

- Automated creation of Normal Behavioral Model (NBM) pipelines.
- Attention added to LSTM cells for better time-series modeling.
- Introduction of the Darwin Run-Time Engine (RTE), which provides ability to download models and run them outside the Darwin cloud.
- New routes created:
  - Cleaning a dataset
  - Downloading a dataset, whether it is a cleaned dataset or the original dataset
  - Downloading a model
- More explicit model details provided during *lookup_model* call.
- Improved error logging when differences between the training and testing datasets are found.
- Feature engineering has been improved and is much faster.

Fixed Issues in 1.5

- Properly return partial results from backpropagation when time runs out during supervised modeling.
- Additional error handling in data profiler.
- Can now handle imbalanced batches for unsupervised models.
- Fixed an issue where the remaining run time was not updated properly after the initial backpropagation step during supervised modeling.
- Fixed a segfault when predictions were run on a cpu.
- Properly pass in validation set when using multiprocessing for backprop distribution.
- Spaces and some special characters can now be used in dataset and model names.
- Fixed an issue where upgrading Keras to version 2.2.0 would break unsupervised.
- Fixed an issue where TerminateHandler returns NoneType with a loss of NaN or inf.

Known Issues in 1.5

- Models created from earlier versions of Darwin are incompatible with version 1.5. These models need to be re-created.
- Re-training or resuming training on a model should be done with the original dataset, since a different dataset may not have the same categories for each feature as the original dataset.
- Data submitted to *run_model* must have the same number of columns and column headers as data submitted to *create_model*, otherwise an error message is returned.
  Note: Affects *create_model, run_model*.
- Setting *recurrent=true* does not work for unsupervised.
  Note: Affects *create_model*.
- Any created models can only specify either *zero* or a single *Target* column.
- Because Darwin cannot one hot encode categorical columns with more than *max_unique_values* in training and test sets, these columns are dropped in test and training sets.
- If the target has more numeric values than the *max_int_unique* set point, the problem is treated as a regression and will use MSE.
- Darwin only drops duplicated columns in data sets with less than 5000 rows.
- Any data set can only have a single (one) date time column or be indexed by date/time, otherwise an error message is returned.
  Note: Affects *create_model, analyze_data*.

General Notes

- Darwin will split the training set into a train and validation set using an 70/30 split:
  - For classification problems, the split will be created using stratified shuffling.
  - For regression problems, the split will be created using random shuffling.
  - For problems with a timestamp (regression or classification problems), no reordering will be done and the last 30% of the input data will be used as validation data. So if sparse time-series data is used for modeling and the important points for predictions are clustered densely together,

there is the potential that the resulting model may only train on non-useful data. If this issue is occurring, try removing the time stamp from the data set.

Contact Support

The following methods enable you to research issues, create a support ticket, or contact SparkCognition:

- Use the Darwin support portal - Read Frequently Asked Questions (FAQ), download documentation, or log your issue.
- **Email Support** - Send email to support@darwinamb.zendesk.com.
- **Phone Support** - The SparkCognition support line is +1-512-956-5576.

Revision Table

Version	Date
v 1.0	02.05.2018
v 1.1	02.22.2018
v 1.2	03.29.2018
v 1.3	05.23.2018
v 1.3.1	06.14.2018
v 1.4	07.31.2018
v 1.5	10.15.2018