SparkCognition™ Darwin™ Release Notes

v 1.3 - 05.23.2018

Darwin Release Notes v 1.3

Darwin release 1.3 has incorporated customer feedback to introduce new approaches to both supervised and unsupervised model generation as well as better explainable AI through the introduction of an *analyze_predictions* endpoint and modification of the *analyze_model* endpoint. The following changes are completed and rolled into the Darwin release 1.3 for immediate use:

New Features

New supervised approach

- Grammar-based approach for building deeper neural networks
- Genetic optimized DNNs with other tree-based models

New unsupervised approach

- Leverage a neural-network-based clustering mechanism that can scale to significantly larger data
- Automated cluster number detection

Change to *analyze_model*

- *analyze_model* now outputs a generalized feature importance for the model.

- If the model is a multi-class model or a clustering model, *analyze_model* can also output a feature importance for each independent cluster or class.

New route: *analyze_predictions*

- Users can now get individualized feature importance per data sample to better explain why certain predictions were made.

Ability to reset password

- Email address is required for user registration
- Existing users can add an email address via the *auth_set_email* route.

Fixed Issues

- You can no longer create a new model with the same name as an existing model. Model names are also now case-sensitive.
- Fixed issue where Target Columns containing non-alphabetic characters can cause problems in API calls.
- Added error message for testing files that do not have header information.

Known Issues

- Darwin will split the training set into a train and validation set using an 80/20 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 20% 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.
- Models created from earlier versions of Darwin are incompatible with version 1.3. These models need to be re-created.
- If a user does not specify a number of clusters for unsupervised, Darwin is likely to select a smaller number to minimize separation space. User should set *n_clusters* to specify a specific number of clusters.
- 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.

- Darwin only drops duplicated columns in data sets with less than 5000 rows.
- When using *feature_eng*, the percent complete reads *0%* until the automated windowing, preprocessing, and feature engineering is complete.

  Notes:

  - Depending on data set size, this process can require substantial time to complete.
  - Affects *create_model*.
- Any dataset 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*.
- *max_int_uniques* keyword parameter must be a natural number, otherwise an error message is returned.
  Note: Affects *create_model*, *analyze_data*.
- *max_unique_values* keyword parameter must be a natural number, otherwise an error message is returned.
  Note: Affects *create_model*, *analyze_data*.
- *window_len* parameter (optional) to *create_model* must:

  1. be a natural number,
  2. be smaller than the size of any training set, and
  3. be smaller than the size of any test set, otherwise an error message is returned.

  Note: Affects *create_model*, *analyze_data*.
- *feature_select* parameter must be a real number in [0,1], otherwise an error message is returned.
  Note: Affects *create_model*, *analyze_data*.

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