



SparkCognition™ Darwin™ Release Notes

v 1.6.1 - 02.06.2019

This document contains copyrighted and proprietary information of SparkCognition and is protected by United States copyright laws and international treaty provisions. No part of this publication may be reproduced, stored in a retrieval system or transmitted in any form or by any means, electronic, mechanical, photocopying, recording, scanning or otherwise, except as permitted under such laws or with the prior written permission of SparkCognition Inc.

SparkCognition™, the sparkcognition logo, Darwin™, DeepArmor®, DeepNLP™, MindFabric®, SparkSecure® and SparkPredict™, are trademarks of SparkCognition, Inc. and/or its affiliates and may not be used without written permission. All other trademarks are the property of their respective owners.

©SparkCognition, Inc. 2017-2019. All rights reserved.

Darwin Release Notes v 1.6.1

Darwin release 1.6.1 is a hotfix release targeted to defects that were fixed. We have included the new features section from the 1.6 release for your convenience. The following fixed issues are completed and Darwin release 1.6.1 is ready for immediate use:

New Features in 1.6

- Model building using larger datasets. The maximum file size is now 500 MB for unsupervised and Normal Behavioral Modeling (NBM) and 10 GB for supervised. As a result of this support, the cleaning of data step has been separated from the model creation step, which means that you must clean your data prior to model building.
- Export models to either JSON or ONNX format. **Note:** ONNX is only supported for supervised Deepnet models. Exporting to ONNX is not currently available through the UI.
- Users can view and download the Elites of the model population (top Deepnet, top Random Forest, or top XGBoost model)
- Time-series models completed in version 1.6.x are more accurate than in previous versions due to the addition of temporal convolutional networks to the architecture search
- Data can be run successfully through downloaded models and inserted into the Darwin Runtime Engine (RTE)

- NBM models can be downloaded and run in the Darwin RTE
- Users can specify the Error function they care about fitting to when creating a model
- IsolationForest Outlier detection has been added for unsupervised anomaly detection
- A minimum recommended training time is now calculated based on an input dataset
- Users can force Darwin to treat certain problems as regression even if there are limited samples

Fixed Issues in 1.6.1

- Fixed an issue where you could not clean test data through the small data pipeline against a model that was created by a dataset cleaned by the big data pipeline
- The Darwin RTE was throwing errors during some preprocessing changes for large data models
- The Darwin RTE was not running predictions on large data models
- Evaluate jobs were stuck in a requested state
- Evaluate call was failing to download the entity

Known Issues in 1.6.1

- The Darwin RTE does not support unsupervised models. It only supports supervised and NBM models.
- Analyze predictions is not supported for big datasets (> 500 MB).
- Non-target columns that begin with the same prefix as the target are considered to be the target. To avoid this issue, ensure that the columns do not begin with the same prefix as the target. For example, if your target column is "temperature", ensure that no other columns begin with the string "temperature". A column named "temperature_change" would be interpreted as the target and cause this issue.
- When large datasets are cleansed, they are divided into 1 GB parts. There is an issue where the download dataset function will only download part 1 of a multi-part cleansed dataset. This issue will be fixed in the next release. This does not affect the training or running of models using large datasets.
- 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*.

General Notes

- 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.

- 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.
- 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*.

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
v 1.6	01.16.2019
v 1.6.1	02.06.2019