SparkCognition™ Darwin™ Release Notes

v 2.0.2 - 09.04.2019, updated 09.12.2019

Darwin Release Notes v 2.0.2

The documentation for this version of Darwin includes:

- The *Darwin Release Notes*, version 2.0.2
- The *Darwin User Interface Guide*, version 2.0.2
- The *Darwin API User Guide*, version 1.34.1
- The *Darwin Python SDK User Guide*, version 1.44.0

All of these documents are available for download from the Darwin support portal.

Darwin release 2.0.2 is focused on fixing issues and improving the overall user experience. Version 2.0.1 was an internal use only release that did not have any end-user functionality.

New Features introduced in version 2.0 and 2.0.2

- New data analysis and recommendation service providing cleaning and analysis of datasets as they are uploaded as well as a guided UI workflow for creating models
- New Forecasting Models accessible through the SDK
- Manual and Automatic Train/Validation splitting
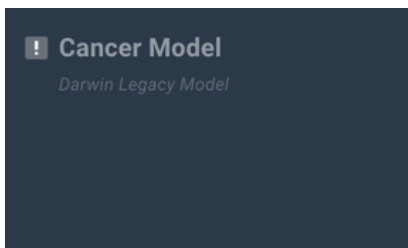- Overfitting improvements

- Support links and a feedback form available from the UI
- Improved UI Error messaging
- UTC time is used for SDK/API; local time is used for UI
- (New in 2.0.2) Multi-select of datasets in the Data Bank
- (New in 2.0.2) Improvements to the Data Explorer functionality
- (New in 2.0.2) Improvements to data ingestion to detect malformed .csv files

Fixed Issues in 2.0.2

- Fixed an issue where display population was not working for on-premise instances
- The Japanese version no longer has time stamps displayed in English
- Fixed an issue where columns were excluded for too many categories even though it was changed to numeric
- Fixed an issue where a low variance warning should be a no variance display
- Fixed an issue where cleaning a dataset with no target column produced an error
- Fixed a scrolling issue in the Data Viewer

Known Issues in 2.0.2

- Models created in version 1.x are not compatible with version 2.0.x. Those models will display an error similar to the following and will need to be re-created.

- Date/Time columns are currently dropped. If you want to create a time series problem, you must perform one of the following:

    - In the SDK, set `recurrent=True`, or
    - In the UI, select `Yes` on the Time Series problem page.

  In both of these cases, Darwin will automatically perform nested cross-validation, use forward fill imputation, and turn on recurrent architectures to treat the problem as a time series problem.

- The maximum number of features that can be displayed in the UI is 350. For best usability, ensure that your datasets have less than 350 features.

- Date/Time columns will be displayed incorrectly in the Model Results page. In these cases, the date columns will be shown as categorical.

- The Mean Squared Error (MSE) loss value presented in the UI is the scaled MSE loss, whereas the MSE presented in the Model Results view is the unscaled MSE loss.

- When uploading a dataset, ensure that you wait until the upload is complete. Other UI functionality may not be available while the dataset upload is in progress.

- When downloading large datasets, you are not notified until the download is complete. Do not log out of Darwin until the download is complete. (Download times may vary depending on bandwidth.)

- Occasionally, the Loss and Algorithm are not populated in the UI. This does not necessarily mean that the model failed to build, and upon clicking the model card, you should still see the training performance results and be able to run the model to produce predictions.

- When downloading an artifact using the Runtime Engine (RTE), it is not being downloaded to the user-defined path. The RTE is saving the artifact in a temporary folder on the local machine. The download confirmation will output the temp folder path.

- When exporting a model, the ONNX format is only available for neural network models. The JSON format is available for all model downloads, including neural network models.

- The Darwin RTE does not support unsupervised models nor models with TCN architectures. It only supports supervised and NBM models.

- Analyze predictions is not supported for large datasets (> 500 MB).

General Notes

- 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 (set to 30), these columns are dropped in test and training sets.
- Darwin only drops duplicated columns in data sets with less than 5000 rows.
- Users must now set `recurrent = True` with the SDK and API in order to see the LSTM and TCN models used.
- The Darwin UI may not show data when header names contain more than 63 characters. Model building will operate as expected, but components of the visualization may not appear as designed.
- Odd behavior might be encounted if you try to predict categories on a numeric target or predict values on a categorical target. Ensure that the type of prediction you want is in alignment with the type of target.
- Any data set can only have a single (one) date time column or be indexed by date/time, otherwise an error message is returned.

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 darwin_support@sparkcognition.com.
- **Phone Support** - The SparkCognition support line is +1-512-400-2001.

Revision Table

Version	Date
v 1.6	01.16.2019
v 1.6.1	02.06.2019
v 1.6.2	03.25.2019
v 1.7	05.16.2019
v 2.0	07.29.2019
v 2.0.1	Internal use only
v 2.0.2	09.04.2019