infor

Cheat sheet

Infor OS Coleman ML

📋 Key concepts & definitions

Coleman ML	Infor's machine learning platform.
Quest	A flow of activities that make up the machine learning model.
Training Quest	A quest involving a predictive method that produces a trained model.
Trained Model	A model that can be used to predict outcomes based on new data.
Production Quest	A trained model that has taken steps to deployment for access via endpoints.
Endpoint	The REST API access point of a production quest to process new data through the model. Endpoints can process data by being passed a CSV, a JSON message, or accessed via the ION API gateway.
Data lake	Flexible and economical cloud object storage solution where data is stored in its raw format. This is Coleman ML's primary data source
Label/Target	Terms that refer to the predicted variable of the model.
Categorical	Data types that are non-numeric in nature and belong to a category instead. E.g. "Country of Residence".
Supervised Learning	Machine learning algorithms that form relationships between targeted label and input features so that the output values for unseen data can be predicted. Supervised algorithms must be trained on known outcomes.
Unsupervised Learning	Machine learning algorithms that make inferences from data using only input features without referring to known or labelled outcomes. These algorithms can discover data structures by clustering it into intuitive groups.

🔗 Algorithms

XGBoost	XGBoost algorithm (supervised), or Extreme Gradient Boosting is a decision-tree-based algorithm used for classification and regression problems. It trains and predicts with many models parallelly to produce a single superior output. Moreover, it uses a gradient descent algorithm to minimize the loss when adding new models.
Random Forest	Random Forest algorithm (supervised) is a decision-tree-based algorithm used for binary, multi-class classification or regression problems. It constructs and combines multiple decision trees to provide a more accurate prediction. Different from Decision Tree, the Random Forest algorithm randomly selects observations and features, and builds several decision trees before averaging the results.
Multilayer Perceptron	Multilayer Perceptron algorithm, or MLP (supervised) is a class of the feedforward artificial neural networks. It can be used both for classification and regression problems. A MLP consists of at least three layers of nodes: an input layer, a hidden layer and an output layer. Except for the input nodes, each node is a neuron that uses a nonlinear activation function. It can distinguish data that is not linearly separable.
Deep AR	DeepAR Forecasting algorithm (supervised) is a learning algorithm for forecasting scalar (one-dimensional) time series using recurrent neural networks (RNN). It trains a model by randomly sampling several training examples from each of the time series in the training dataset. You can use the trained model to generate forecasts for new time series that are similar to the ones it has been trained on.
Principal Component Analysis	Principal Component Analysis algorithm (unsupervised), or PCA, is a dimensionality-reduction algorithm used to reduce the dimensionality of large data sets, by transforming a large set of variables into a smaller one that still contains most of the information from the large set.
Custom Algorithm	Allows for the implementation of your own custom algorithm instead of one of the current algorithm tools.

📋 Resources

Product overview	YouTube playlist	Documentation	Technology blog	User community

Infor OS Coleman APIs

v1/endpoints	GET /v1/endpoints	Get a list of endpoints with in-service status.
	POST /v1/endpoints/{endpoint}/detailed-prediction	Retrieve a detailed prediction for a given endpoint.
	POST /v1/endpoints/{endpoint}/prediction	Retrieve a prediction for a given endpoint.
	GET /v1/endpoints/{endpoint}/schema	Get the schema for a given endpoint.

Quest flow

Datasets	Each quest starts with a dataset. These datasets can come from static files, or data lake imports.
Prepare Data	Prepare data blocks accomplish data modification tasks to get data from various sources to a structure that the machine learning model can learn/predict from.
Explore Data	Explore data blocks allow for the creation of plots, charts, and heatmaps useful for understanding data relationships.
Apply Algorithm	Apply Algorithm blocks contain basic machine learning algorithms for use, or allow for custom algorithms to be implemented.
Train Model	Train Model blocks ingest a training dataset and the chosen algorithm type to train.
Evaluate Model	Evaluate Model blocks provide performance statistics against a test set, or compare performance of multiple models against each other.
Realtime Production	Real time production quests will not train a new model, but take the trained model and apply it to new data. Trained quests must be deployed for real time production to be accessed outside of Coleman
Batch Production	Batch production quests deploy a trained model as a data pipeline for batch transformation jobs. Intended for use cases working with larger datasets which consume and ingest to the Data Lake.
Endpoint	A Realtime Production quest can be deployed as an endpoint accessible by the ION API Gateway.

Sample quest