Machine Learning Toolkit

Use this document for a quick list of ML search commands as well as some tips on the more widely used algorithms from the Machine Learning Toolkit.

Search Commands for Machine Learning	The Machine Learning Toolkit provides custom search commands for applying machine learning to your data.	
Command	**Description**	**Syntax**
fit	Fit and apply a machine learning model to search results.	`... | fit algorithm y from x params into model_name as output_field`
apply	Apply a machine learning model that was learned using the fit command.	`... | apply model_name as output_field`
summary	Return a summary of a machine learning model that was learned using the fit command.	`| summary model_name`
listmodels	Return a list of machine learning models that were learned using the fit command.	`| listmodels`
deletemodel	Delete a machine learning model that was learned using the fit command.	`| deletemodel model_name`
sample	Randomly sample or partition events.	`... | sample options by split_by_field`
score	Run statistical tests to validate model outcomes.	`... | score method actual predicted options`

FREQUENTLY USED ALGORITHMS

Anomaly Detection	Find events that contain unusual combinations of values.
Algorithm	**Examples**
LocalOutlierFactor	`... | fit LocalOutlierFactor * n_neighbors=10 algorithm=kd_tree metric=minkowski p=1 contamination=0.14 leaf_size=10`
OneClassSVM	`... | fit OneClassSVM * kernel=poly nu=0.5 coef0=0.5 gamma=0.5 tol=1 degree=3 shrinking=f into TESTMODEL_OneClassSVM`

Feature Extraction	Feature extraction algorithms transform fields for better prediction accuracy.
Algorithm	**Examples**
FieldSelector	`... | fit FieldSelector type=categorical SLA_violation from *`
KernelPCA	`... | fit KernelPCA * k=3 gamma=0.001`
PCA	`... | fit PCA * k=3`
TFIDF	`... | fit TFIDF Reviews into user_feedback_model max_def=0.6 min_def=0.2`

Preprocessing	Preprocessing algorithms are used for preparing data and help with prediction accuracy.
Algorithm	**Examples**
RobustScaler	`... | fit RobustScaler *`
StandardScaler	`... | fit StandardScaler *`

Cluster Numeric	Partition events with multiple numeric fields into clusters.
Algorithm	**Examples**
Birch	`... | fit Birch * k=3`
DBSCAN	`... | fit DBSCAN * eps=0.9`
KMeans	`... | fit KMeans * k=3`
SpectralClustering	`... | fit SpectralClustering * k=3`
XMeans	`... | fit XMeans *`

Forecasting	Forecast future values given past values of a metric (numeric time series).
Algorithm	**Examples**
ARIMA	`... | fit ARIMA Voltage order=4-0-1`

Predict Numeric

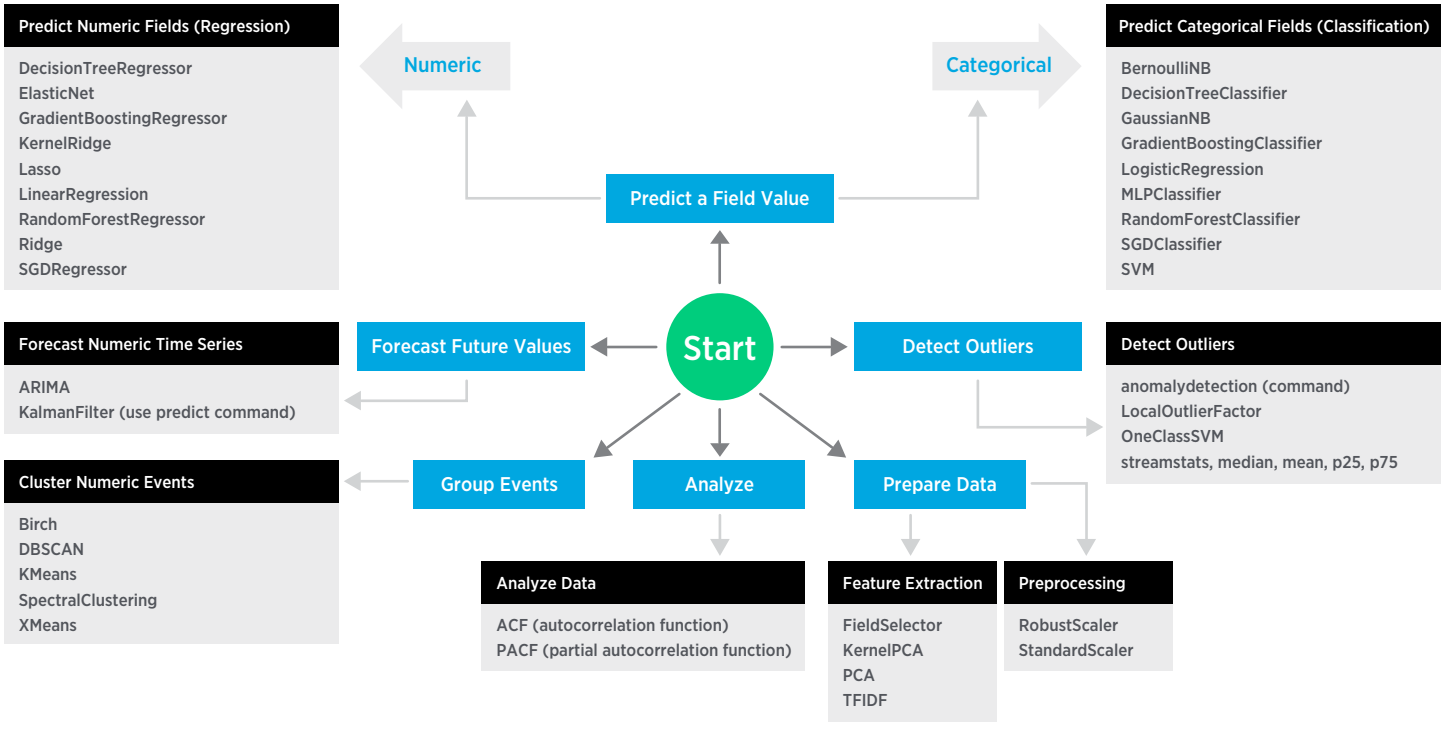
Predict the value of a numeric field using the values of other fields in that event.

Algorithm	Examples
DecisionTreeRegressor	... | fit DecisionTreeRegressor temperature from date_month date_hour into temperature_model
ElasticNet	... | fit ElasticNet temperature from date_month date_hour normalize=true alpha=0.5 into temperature_model
GradientBoostingRegressor	... | fit GradientBoostingRegressor temperature from date_month date_hour into temperature_model
KernelRidge	... | fit KernelRidge temperature from date_month date_hour into temperature_model
Lasso	... | fit Lasso temperature from date_month date_hour into temperature_model
LinearRegression	... | fit LinearRegression temperature from date_month date_hour into temperature_model
RandomForestRegressor	... | fit RandomForestRegressor temperature from date_month date_hour into temperature_model
Ridge	... | fit Ridge temperature from date_month date_hour normalize=true alpha=0.5 into temperature_model
SGDRegressor	... | fit SGDRegressor temperature from date_month date_hour into temperature_model

Predict Categorical

Predict the value of a categorical field using the values of other fields in that event.

Algorithm	Examples
BernoulliNB	... | fit BernoulliNB species from * alpha=0.5 binarize=0 fit prior=f into species_model
DecisionTreeClassifier	... | fit DecisionTreeClassifier SLA_violation from * into sla_model
GaussianNB	... | fit GaussianNB species from * into species_model
GradientBoostingClassifier	... | fit GradientBoostingClassifier species from * into species_model
LogisticRegression	... | fit LogisticRegression SLA_violation from IO_wait_time into sla_model
MLPClassifier	... | fit MLPClassifier species from * into species_model
RandomForestClassifier	... | fit RandomForestClassifier SLA_violation from * into sla_model
SGDClassifier	... | fit SGDClassifier SLA_violation from * into sla_model
SVM	... | fit SVM SLA_violation from * into sla_model



Download Machine Learning Toolkit. Read the Machine Learning Documentation.

splunk> docs.splunk.com