cyber & data

"From bits to information"

Introduction to Splunk and Machine Learning – Part 1

Outline

- Introduction to Machine Learning.
- Introduction to model fitting.

Machine Learning

The two main approaches used within the detection of threats is within:

- **Signature detection**, where we match against well-known patterns of malicious behavior.
- **Use anomaly detection**, where we define a normal behavior pattern.

Within a decision engine, we often use the concept of correct guesses (true) and incorrect ones (false). So, a true positive is where we determine that an event was correctly detected, while a false positive is where a true event was not detected (and thus missed by the system).

Within IDS (Intrusion Detection Systems) there is often a balance to be struck when tuning the systems so that users do not get swamped by too many false alerts (false positives), or from too many fake alerts.

Machine Learning

- **Information sources**. This involves defining the sources of information that would be required to capture the right information.
- **Data capturing tools**. This involves creating the software agents required to the required data.
- **Data pre-processing**. This involves processing the data into a format which is ready for the analysis part.
- **Feature extraction**. This involves defines the key features that would be required to the analysis engine.
- **Analysis engine**. This involves the creation of an analysis engine which takes the features and creates scoring to evaluate risks.
- **Decision engine**. This takes the scoring systems from the analysis stage and makes a reasoned decision on the level of risk involved.

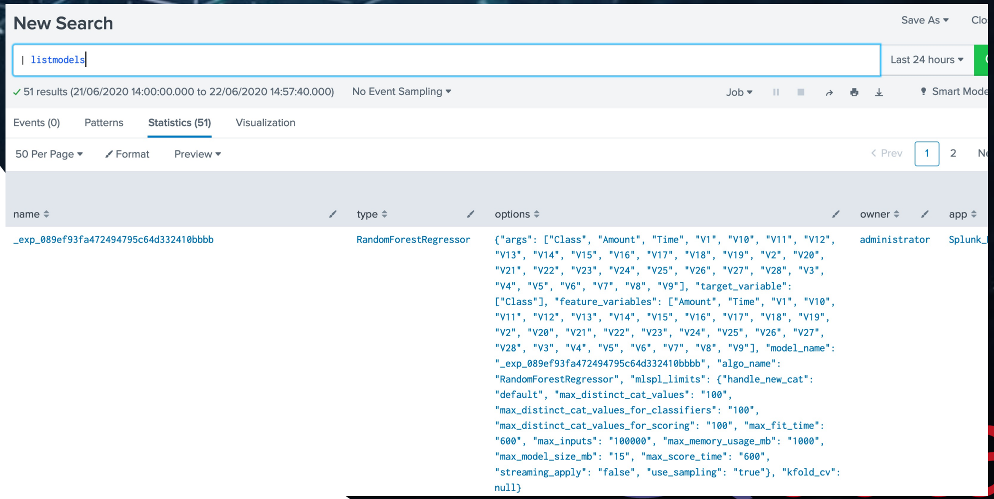
Splunk and Machine Learning

- fit. Fit a model
- apply. Apply a model run by the fit command.
- summary. Show summary of model.
- listmodels. List the models.
- deletemodel. Delete a model.
- score. Show scores for tests.

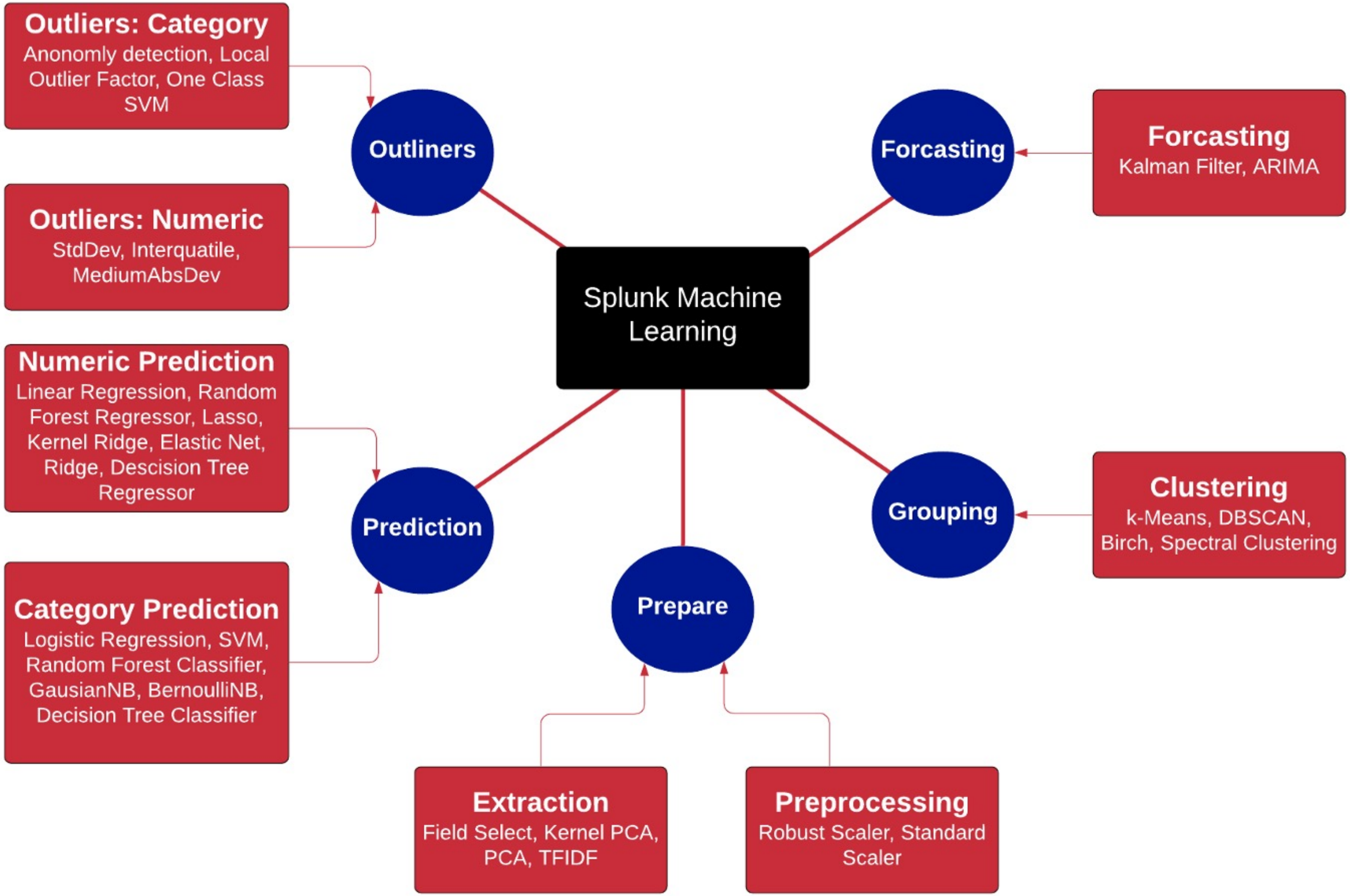
| inputlookup iris.csv
| fit GaussianNB petal_length from * into myModel
| apply myModel as new_petal link.

| summary myModel link

| listmodels link



Machine Learning

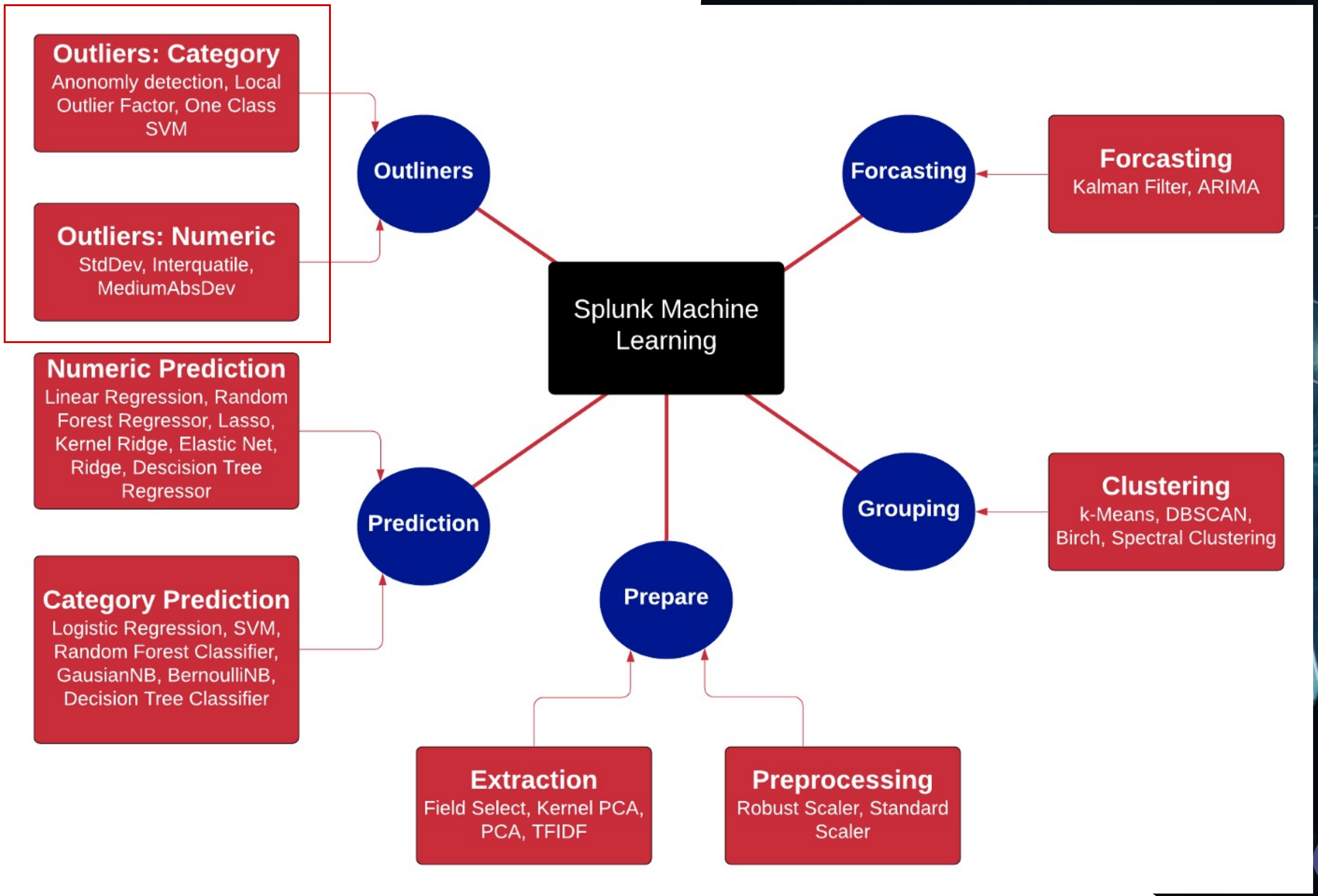


cyber & data

“From bits to information”

fit command

Outliners



cyber & data

fit- Anomaly Detection

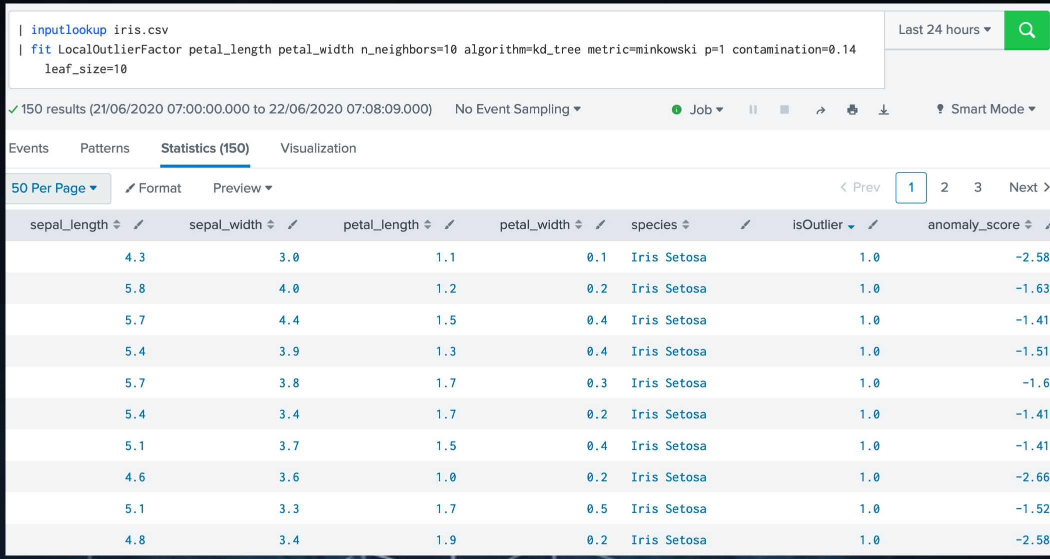
| inputlookup iris.csv
| fit **LocalOutlierFactor** petal_length petal_width n_neighbors=10 algorithm=kd_tree metric=minkowski p=1 contamination=0.14 leaf_size=10 Link.

- *class* sklearn.neighbors.LocalOutlierFactor(*n_neighbors=20, *, algorithm='auto', leaf_size=30, metric='minkowski', p=2, metric_params=None, contamination='auto', novelty=False, n_jobs=None* Link. -> Local Outlier Factor.

| inputlookup iris.csv
| fit **OneClassSVM** * kernel="poly" nu=0.5 coef0=0.5 gamma=0.5 tol=1 degree=3 shrinking=f into TESTMODEL_OneClassSVM Link.

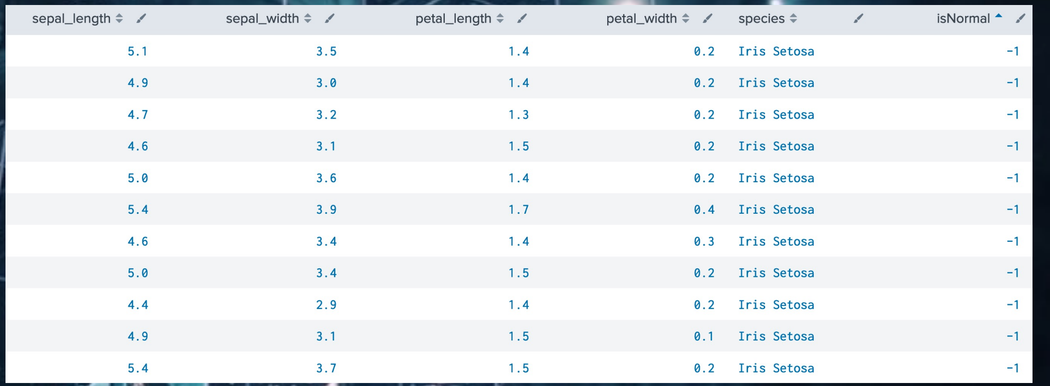
- *class* sklearn.svm.OneClassSVM(*, *kernel='rbf', degree=3, gamma='scale', coef0=0.0, tol=0.001, nu=0.5, shrinking=True, cache_size=200, verbose=False, max_iter=-1*) Link

| inputlookup call_center.csv
| fit **DensityFunction** count by "source" into mymodel Link.



```
| inputlookup iris.csv
| fit LocalOutlierFactor petal_length petal_width n_neighbors=10 algorithm=kd_tree metric=minkowski p=1 contamination=0.14
  leaf_size=10
```

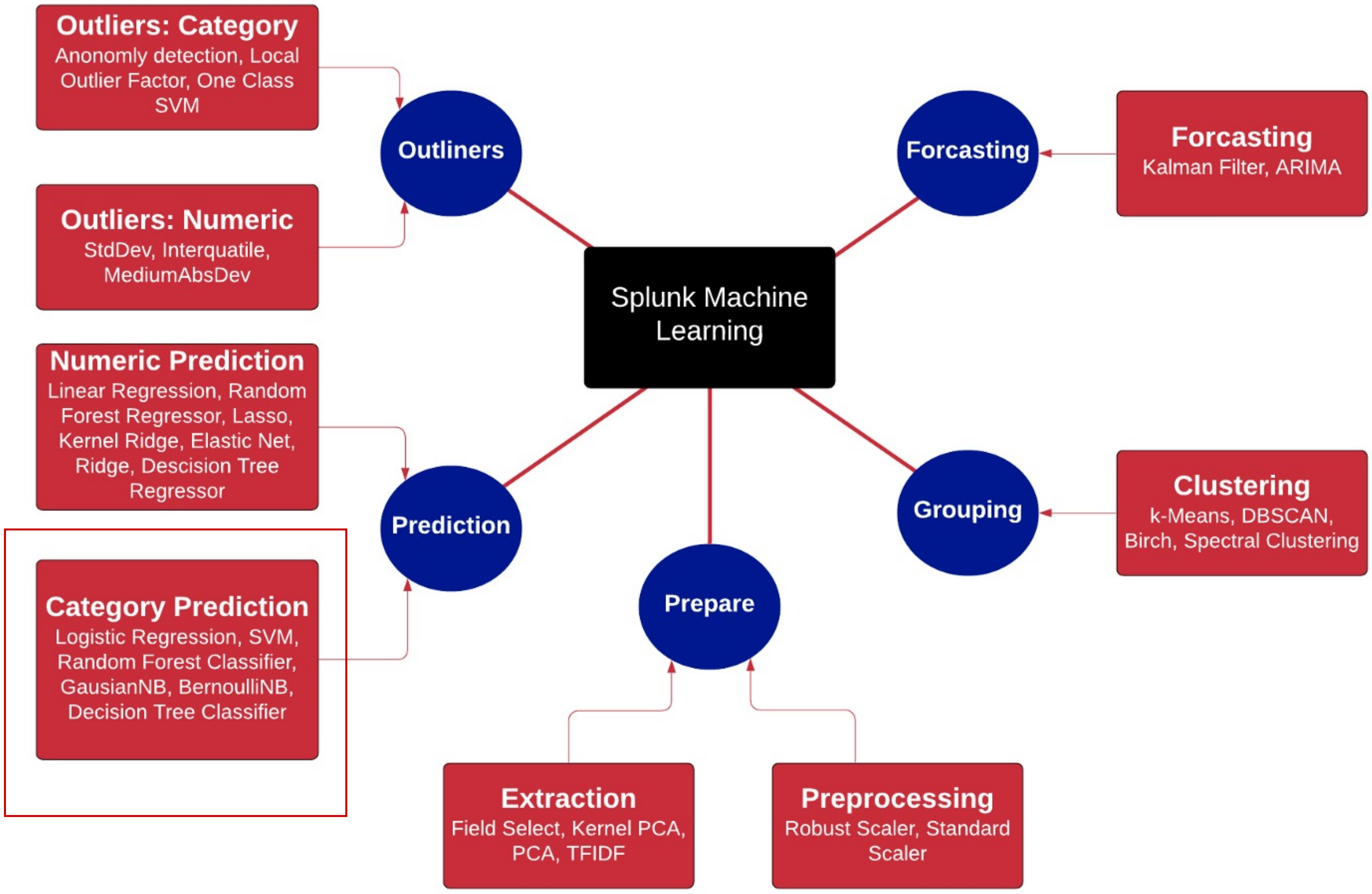

sepal_length	sepal_width	petal_length	petal_width	species	isOutlier	anomaly_score
4.3	3.0	1.1	0.1	Iris Setosa	1.0	-2.58
5.8	4.0	1.2	0.2	Iris Setosa	1.0	-1.63
5.7	4.4	1.5	0.4	Iris Setosa	1.0	-1.41
5.4	3.9	1.3	0.4	Iris Setosa	1.0	-1.51
5.7	3.8	1.7	0.3	Iris Setosa	1.0	-1.6
5.4	3.4	1.7	0.2	Iris Setosa	1.0	-1.41
5.1	3.7	1.5	0.4	Iris Setosa	1.0	-1.41
4.6	3.6	1.0	0.2	Iris Setosa	1.0	-2.66
5.1	3.3	1.7	0.5	Iris Setosa	1.0	-1.52
4.8	3.4	1.9	0.2	Iris Setosa	1.0	-2.58



sepal_length	sepal_width	petal_length	petal_width	species	isNormal
5.1	3.5	1.4	0.2	Iris Setosa	-1
4.9	3.0	1.4	0.2	Iris Setosa	-1
4.7	3.2	1.3	0.2	Iris Setosa	-1
4.6	3.1	1.5	0.2	Iris Setosa	-1
5.0	3.6	1.4	0.2	Iris Setosa	-1
5.4	3.9	1.7	0.4	Iris Setosa	-1
4.6	3.4	1.4	0.3	Iris Setosa	-1
5.0	3.4	1.5	0.2	Iris Setosa	-1
4.4	2.9	1.4	0.2	Iris Setosa	-1
4.9	3.1	1.5	0.1	Iris Setosa	-1
5.4	3.7	1.5	0.2	Iris Setosa	-1

source	_time	count	IsOutlier(count)	BoundaryRanges
si_active_agents	2017-09-01 00:00	1	0.0	-Infinity:1.0:0.005 1.0:Infinity:0.005
si_active_agents	2017-09-01 01:00	1	0.0	-Infinity:1.0:0.005 1.0:Infinity:0.005
si_active_agents	2017-09-01 02:00	1	0.0	-Infinity:1.0:0.005 1.0:Infinity:0.005
si_active_agents	2017-09-01 03:00	1	0.0	-Infinity:1.0:0.005 1.0:Infinity:0.005
si_active_agents	2017-09-01 04:00	1	0.0	-Infinity:1.0:0.005 1.0:Infinity:0.005

Prediction (Cat)



cyber & data

fit- Prediction

| inputlookup iris.csv
| fit **AutoPrediction** random_state=42 petal_length from * max_features=0.1 into auto_classify_model test_split_ratio=0.3 random_state=42 Link.

| inputlookup iris.csv
| fit **BernoulliNB** petal_length from * into TESTMODEL_BernoulliNB alpha=0.5 binarize=0 fit_prior=f Link.

- *class* sklearn.naive_bayes.BernoulliNB(**, alpha=1.0, binarize=0.0, fit_prior=True, class_prior=None*) Link.

| inputlookup iris.csv
| fit **DecisionTreeClassifier** petal_length from * into sla_ MOD Link.

| inputlookup iris.csv
| fit **GaussianNB** petal_length from * into MOD Link.

| inputlookup iris.csv
| fit **GradientBoostingClassifier** petal_length from * into MOD link

sepal_length	sepal_width	petal_length	petal_width	species	predicted(petal_length)
5.1	3.5	1.4	0.2	Iris Setosa	1.4200000000000002
4.9	3.0	1.4	0.2	Iris Setosa	1.73
4.7	3.2	1.3	0.2	Iris Setosa	1.3400000000000003
4.6	3.1	1.5	0.2	Iris Setosa	1.4700000000000002
5.0	3.6	1.4	0.2	Iris Setosa	1.4200000000000002
5.4	3.9	1.7	0.4	Iris Setosa	1.6600000000000001
4.6	3.4	1.4	0.3	Iris Setosa	1.44
5.0	3.4	1.5	0.2	Iris Setosa	1.4700000000000002

sepal_length	sepal_width	petal_length	petal_width	species	predicted(petal_length)
5.1	3.5	1.4	0.2	Iris Setosa	1.2
4.9	3.0	1.4	0.2	Iris Setosa	1.2
4.7	3.2	1.3	0.2	Iris Setosa	1.2
4.6	3.1	1.5	0.2	Iris Setosa	1.2
5.0	3.6	1.4	0.2	Iris Setosa	1.2
5.4	3.9	1.7	0.4	Iris Setosa	1.7
4.6	3.4	1.4	0.3	Iris Setosa	1.4
5.0	3.4	1.5	0.2	Iris Setosa	1.2

&cyber data

fit- Prediction

| inputlookup iris.csv
| fit **LogisticRegression** petal_length from * into MOD Link.

| inputlookup iris.csv
| fit **MLPClassifier** petal_length from * into MOD Link.

| inputlookup iris.csv
| fit **RandomForestClassifier** petal_length from * into MOD Link.

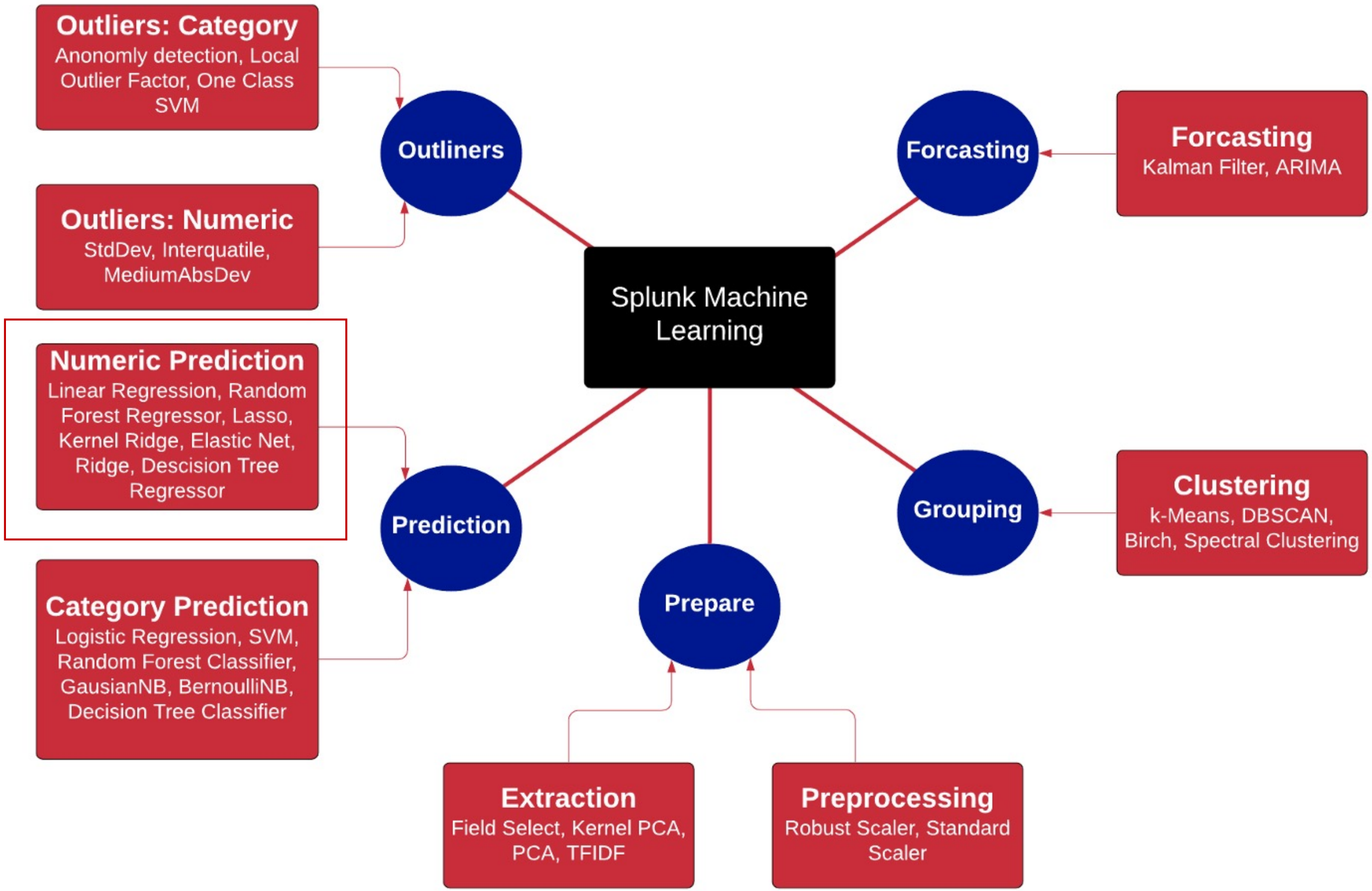
| inputlookup iris.csv
| fit **SGDClassifier** petal_length from * into MOD Link.

| inputlookup iris.csv
| fit **SVM** petal_length from * into MOD. Link.

sepal_length	sepal_width	petal_length	petal_width	species	predicted(petal_length)
5.1	3.5	1.4	0.2	Iris Setosa	1.4200000000000002
4.9	3.0	1.4	0.2	Iris Setosa	1.73
4.7	3.2	1.3	0.2	Iris Setosa	1.3400000000000003
4.6	3.1	1.5	0.2	Iris Setosa	1.4700000000000002
5.0	3.6	1.4	0.2	Iris Setosa	1.4200000000000002
5.4	3.9	1.7	0.4	Iris Setosa	1.6600000000000001
4.6	3.4	1.4	0.3	Iris Setosa	1.44
5.0	3.4	1.5	0.2	Iris Setosa	1.4700000000000002

sepal_length	sepal_width	petal_length	petal_width	species	predicted(petal_length)
5.1	3.5	1.4	0.2	Iris Setosa	1.2
4.9	3.0	1.4	0.2	Iris Setosa	1.2
4.7	3.2	1.3	0.2	Iris Setosa	1.2
4.6	3.1	1.5	0.2	Iris Setosa	1.2
5.0	3.6	1.4	0.2	Iris Setosa	1.2
5.4	3.9	1.7	0.4	Iris Setosa	1.7
4.6	3.4	1.4	0.3	Iris Setosa	1.4
5.0	3.4	1.5	0.2	Iris Setosa	1.2

Prediction (Numeric)



fit- Prediction

| inputlookup track_day_missing.csv
| fit **AutoPrediction** batteryVoltage target_type=numeric test_split_ratio=0.7 from * into PM Link.

vehicleType	batteryVoltage	engineCoolantTemperature	engineSpeed	lateralGForce	longitudeGForce	speed	verticalGForce	predicted(batteryVoltage)
2015 Porsche GT3		93.0	6060	1.11	0.5	69	-2.0	13.90133565917362
2013 Audi RS5	13.937000000000001	94.0	4957	0.56	0.7	56	0.95	13.964810194805192
2015 Porsche GT3	13.827	93.0	6163	0.71	0.26	70	-2.0	13.893069210760919
2011 Ford Mustang GT500	14.035	87.0	2846	0.81	-0.71	47	-2.0	14.038379523809521
2015 Porsche GT3	13.827	93.0	6542	0.49	-0.18	76	-2.0	13.965466269841269
2014 Chevrolet Corvette	14.624	105.0	4425	0.32	0.05	100	-0.17	14.70799623015873
2015 Porsche GT3	13.827	93.0	7763	0.24	-0.18	91	-2.0	14.12003358974359

AutoPrediction can support category or numeric, and then calls **RandomForestRegressor**

| inputlookup track_day_missing.csv
| fit **DecisionTreeRegressor** batteryVoltage from * into PM Link.

| inputlookup track_day_missing.csv
| fit **ElasticNet** batteryVoltage from * into EN. Link.

ElasticNet is a linear regression model and is a generalised form of Lasso and Ridge.

| inputlookup track_day_missing.csv
| fit **GradientBoostingRegressor** batteryVoltage from * into GB Link.

| inputlookup track_day_missing.csv
| fit **KernelRidge** batteryVoltage from * into KR Link.

cyber & data

fit- Prediction

| inputlookup track_day_missing.csv
| fit **Lasso** batteryVoltage from * into LA [Link](#).

| inputlookup track_day_missing.csv
| fit **LinearRegression** batteryVoltage from * into LR [Link](#).

| inputlookup track_day_missing.csv
| fit **RandomForestRegressor** batteryVoltage min_samples_split=30000 from * into RF [Link](#).

| inputlookup track_day_missing.csv
| fit **Ridge** batteryVoltage from * into RD [Link](#).

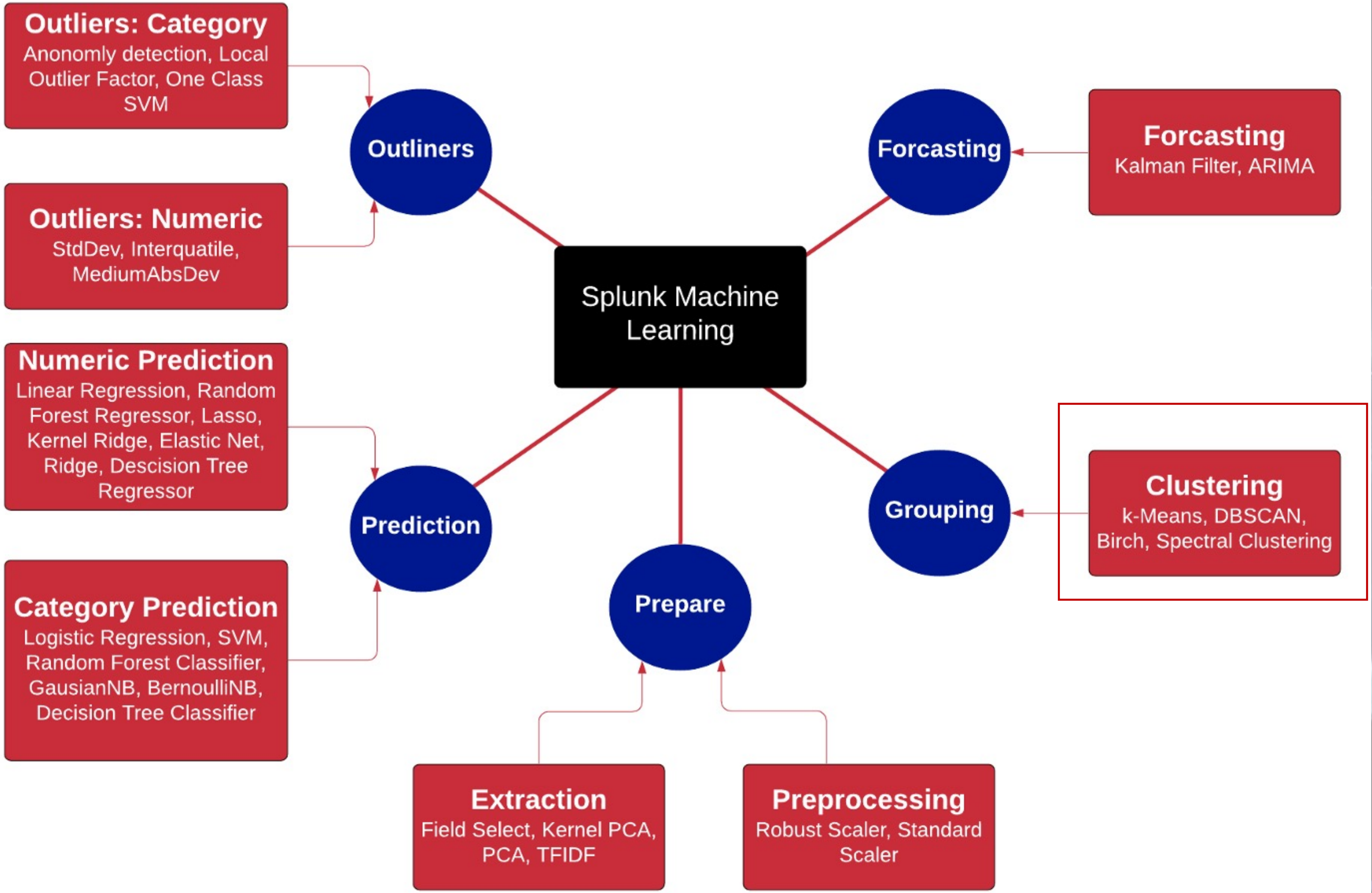
| inputlookup track_day_missing.csv
| fit **SGDRegressor** batteryVoltage from * into SG [Link](#).

| inputlookup app_usage.csv
| fit **SystemIdentification** Expenses from HR1 HR2 ERP dynamics=3-2-2-3 layers=64-64-64 [Link](#).

Uses multi-layer neural network. In this case five layers, with three hidden layers of 64/64/64. Logs are 3 for Expenses, 2 for H1, and so on.

vehicleType	batteryVoltage	engineCoolantTemperature	engineSpeed	lateralGForce	longitudeGForce	speed	verticalGForce	predicted(batteryVoltage)
2015 Porsche GT3		93.0	6060	1.11	0.5	69	-2.0	14.047398311500638
2013 Audi RS5	13.937000000000001	94.0	4957	0.56	0.7	56	0.95	14.050742114165605
2015 Porsche GT3	13.827	93.0	6163	0.71	0.26	70	-2.0	14.047891131013976
2011 Ford Mustang GT500	14.035	87.0	2846	0.81	-0.71	47	-2.0	13.841659798088022
2015 Porsche GT3	13.827	93.0	6542	0.49	-0.18	76	-2.0	14.050547588476304
2014 Chevrolet Corvette	14.624	105.0	4425	0.32	0.05	100	-0.17	14.723515281942252
2015 Porsche GT3	13.827	93.0	7763	0.24	-0.18	91	-2.0	14.052966400552615
2015 Porsche GT3	13.827	93.0	6365	0.38	-0.2	95	-2.0	14.06058143488513
2015 Porsche GT3	13.827	93.0	6713	0.04	0.13	100	-2.0	14.06037488719077

Cluster



cyber & data

fit- Cluster

| inputlookup iris.csv
| fit **Birch** petal_length k=3 partial_fit=true into MOD Link

| inputlookup iris.csv
| fit **DBSCAN** petal_length min_samples=4 link

| inputlookup iris.csv
| fit **GMeans** petal_length random_state=42 into MOD3 link. [based on k-means]

| inputlookup iris.csv
| fit **KMeans** petal_length k=3 into MOD4 link.

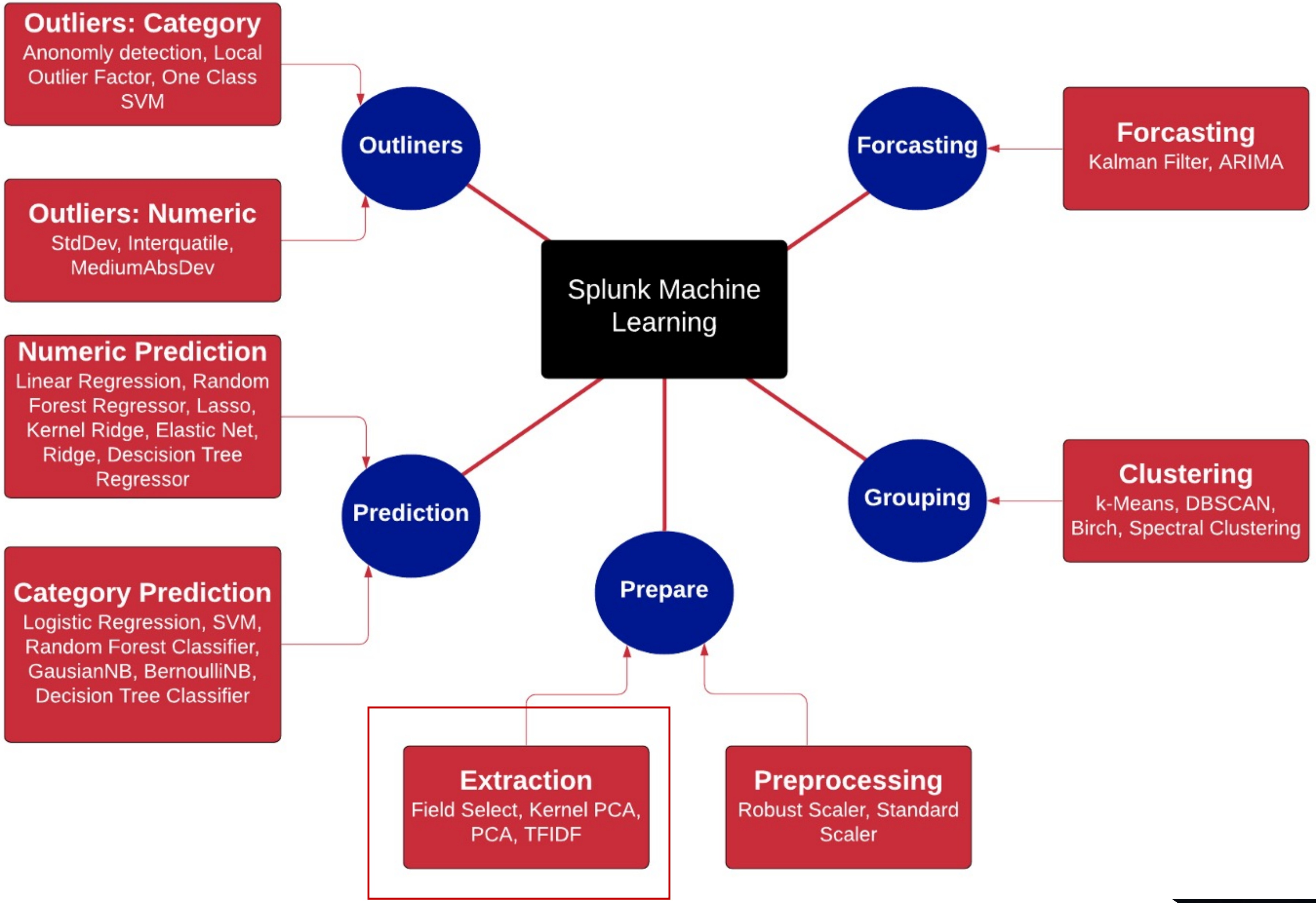
| inputlookup iris.csv
| fit **SpectralClustering** petal_length k=3 link.

| inputlookup iris.csv
| fit XMeans petal_length link.

sepal_length	sepal_width	petal_length	petal_width	species	cluster
7.0	3.2	4.7	1.4	Iris Versicolor	0
6.4	3.2	4.5	1.5	Iris Versicolor	0
6.9	3.1	4.9	1.5	Iris Versicolor	0
5.5	2.3	4.0	1.3	Iris Versicolor	0
6.5	2.8	4.6	1.5	Iris Versicolor	0
5.7	2.8	4.5	1.3	Iris Versicolor	0
6.3	3.3	4.7	1.6	Iris Versicolor	0
4.9	2.4	3.3	1.0	Iris Versicolor	0

sepal_length	sepal_width	petal_length	petal_width	species	cluster	cluster_distance
5.1	3.5	1.4	0.2	Iris Setosa	1	0.004096000000000121
4.9	3.0	1.4	0.2	Iris Setosa	1	0.004096000000000121
4.7	3.2	1.3	0.2	Iris Setosa	1	0.026896000000000267
4.6	3.1	1.5	0.2	Iris Setosa	1	0.0012959999999999383
5.0	3.6	1.4	0.2	Iris Setosa	1	0.004096000000000121
5.4	3.9	1.7	0.4	Iris Setosa	1	0.05569599999999957
4.6	3.4	1.4	0.3	Iris Setosa	1	0.004096000000000121

Feature Extraction



fit- Feature Extraction

| inputlookup track_day.csv
| fit **FieldSelector** batteryVoltage from engineCoolantTemperature, engineSpeed, lateralGForce,longitudeGForce, speed **type=numeric** Link

| inputlookup track_day.csv
| fit **FieldSelector** vehicleType from engineCoolantTemperature, engineSpeed, lateralGForce,longitudeGForce, speed **type=categorical** link

| inputlookup passwords.csv
| fit **HashingVectorizer** Passwords ngram_range=1-2 k=10 Link.

| inputlookup track_day.csv
| fit **ICA** batteryVoltage, engineSpeed n_components=2 as IC Link

| inputlookup track_day.csv
| fit **KernelPCA** batteryVoltage, engineSpeed k=3 gamma=0.001 as PCA Link

vehicleType	batteryVoltage	engineCoolantTemperature	engineSpeed	lateralGForce	longitudeGForce	speed	vertical
2015 Porsche GT3	13.785	93.0	6060	1.11	0.5	69	
2013 Audi RS5	13.937000000000001	94.0	4957	0.56	0.7	56	
2015 Porsche GT3	13.827	93.0	6163	0.71	0.26	70	
2011 Ford Mustang GT500	14.035	87.0	2846	0.81	-0.71	47	

vehicleType	batteryVoltage	engineCoolantTemperature	engineSpeed	lateralGForce	longitudeGForce	speed	verticalGForce	fs_lateralGForce
2015 Porsche GT3	13.785	93.0	6060	1.11	0.5	69	-2.0	1.11
2013 Audi RS5	13.937000000000001	94.0	4957	0.56	0.7	56	0.95	0.56
2015 Porsche GT3	13.827	93.0	6163	0.71	0.26	70	-2.0	0.71
2011 Ford Mustang GT500	14.035	87.0	2846	0.81	-0.71	47	-2.0	0.81

Passwords	Passwords_hashed_0	Passwords_hashed_1	Passwords_hashed_2	Passwords_hashed_3	Passwords_hashed_4	Passwords_ha
qwerty123	-0.0	0.9999999999999998	0.0	-0.0	-0.0	
qwerty5	-1.4719616800160393e-17	0.0	0.015088875493768687	0.4224794055919135	0.7253279855866746	-0.059026953
cisco1	1.0000000000000002	2.2187595626058292e-16	1.0686969971588345e-17	1.2583791481697097e-17	4.44133817001074e-18	-8.1424533116
cisco43	-9.941516562259971e-35	-3.760597526072945e-19	-0.18988975014269943	0.000903986945728833	0.10829789914398458	0.9752121
liverpool	-3.189250306701419e-17	1.9421073798334343e-17	0.48208015086067485	0.34827807416123513	-0.586295299832911	0.17695195
liverp00l	1.4719616800160396e-17	-1.583776923683663e-17	0.7867056762806923	0.13091943941483192	0.29857321546377086	0.10184005
cisco1	1.0	2.2187595626064292e-16	7.027787639449764e-18	-1.6066250926232365e-17	-1.0737492244786327e-17	3.56693015115
cisc01	2.698596413362739e-17	-9.76739240727416e-18	-0.3352717042653099	0.8264821946661363	-0.17112198797126701	-0.06159265
qwerty123	-4.437519125212258e-16	1.0	0.0	-0.0	0.0	

vehicleType	batteryVoltage	engineCoolantTemperature	engineSpeed	lateralGForce	longitudeGForce	sp
2015 Porsche GT3	13.785	93.0	6060	1.11	0.5	
2013 Audi RS5	13.937000000000001	94.0	4957	0.56	0.7	
2015 Porsche GT3	13.827	93.0	6163	0.71	0.26	
2011 Ford Mustang GT500	14.035	87.0	2846	0.81	-0.71	
2015 Porsche GT3	13.827	93.0	6542	0.49	-0.18	76

-2.0 0.000264127291557

vehicleType	batteryVoltage	engineCoolantTemperature	engineSpeed	lateralGForce	longitudeGForce	speed	verticalGForce
2015 Porsche GT3	13.827	93.0	6163	0.71	0.26	70	-2.0
2015 Porsche GT3	13.827	93.0	6542	0.49	-0.18	76	-2.0
2015 Porsche GT3	13.808	92.0	3499	-0.78	-0.42	39	-2.0
2015 Porsche GT3	13.716	93.0	3810	-0.81	-0.46	44	-2.0
2015 Porsche	13.716	93.0	5798	0.21	-0.4	68	-2.0

fit- Feature Extraction

| inputlookup track_day.csv
| fit **NPR** vehicleType from engineSpeed as npr01 [Link]

Normalized Perlich Ratio (NPR) algorithm categorical field values into numeric field entries.

| inputlookup track_day.csv
| fit **PCA** engineCoolantTemperature, engineSpeed, lateralGForce,speed k=3 as pca01 [Link]

Principal Component Analysis (PCA) reduce the number of fields in the data.

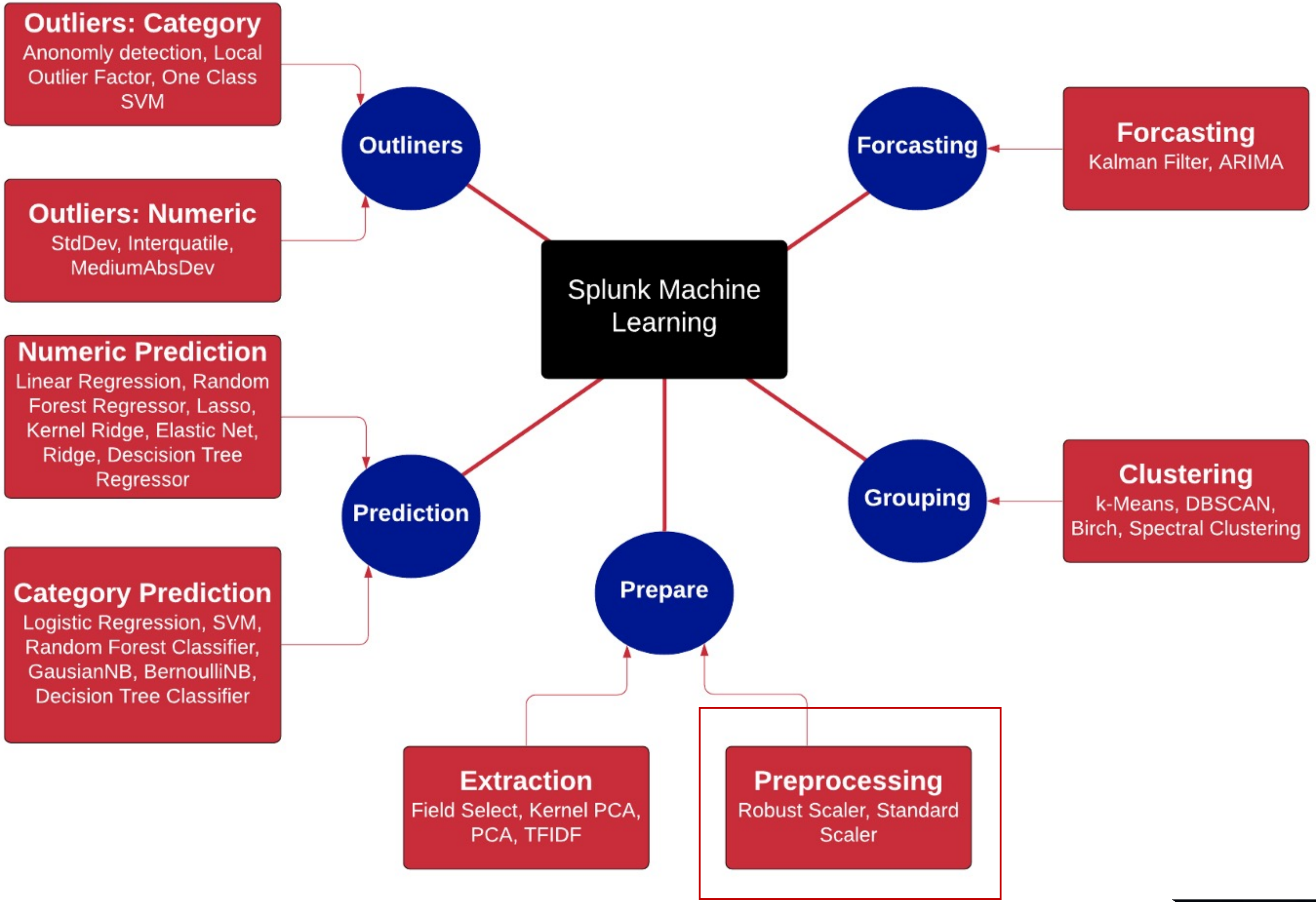
| inputlookup track_day.csv
| fit **TFIDF** vehicleType ngram_range=1-2 max_df=0.6 min_df=0.2 stop_words=english as tf01 [Link]

TFIDF converts raw text data into a matrix.

batteryVoltage	engineCoolantTemperature	engineSpeed	lateralGForce	longitudeGForce	speed	verticalGForce	NPR_engineSpeed_2008 BMW M3	NPR_engineSpeed_2011 Ferrari 458
13.785	93.0	6060	1.11	0.5	69	-2.0	0.20341933065483636	0.0
13.937000000000001	94.0	4957	0.56	0.7	56	0.95	0.4750875723470781	0.0
13.827	93.0	6163	0.71	0.26	70	-2.0	0.15948805818592193	0.0
14.035	87.0	2846	0.81	-0.71	47	-2.0	0.0	0.5529840905437634
13.827	93.0	6542	0.49	-0.18	76	-2.0	0.0	0.0

vehicleType	batteryVoltage	engineCoolantTemperature	engineSpeed	lateralGForce	longitudeGForce	speed	verticalGForce	pca01_1	pca01_2
2015 Porsche GT3	13.785	93.0	6060	1.11	0.5	69	-2.0	1973.31538399881	-3.8336505338179956
2013 Audi RS5	13.937000000000001	94.0	4957	0.56	0.7	56	0.95	870.2387264963066	-3.305800082392769
2015 Porsche GT3	13.827	93.0	6163	0.71	0.26	70	-2.0	2076.3197896193524	-3.8938886305786853
2011 Ford Mustang GT500	14.035	87.0	2846	0.81	-0.71	47	-2.0	-1240.6992929015553	0.3620517152381604
2015 Porsche GT3	13.827	93.0	6542	0.49	-0.18	76	-2.0	2455.3643752526996	-4.943621239329986

Preprocessing



cyber & data

fit- Preprocessing

| inputlookup track_day_missing.csv
| fit **Imputer** batteryVoltage [Link]()

vehicleType	batteryVoltage	engineCoolantTemperature	engineSpeed	lateralGForce	longitudeGForce	speed	verticalGForce	Imputed_batteryVoltage
2015 Porsche GT3		93.0	6060	1.11	0.5	69	-2.0	14.113871297425948
2013 Audi RS5	13.937000000000001	94.0	4957	0.56	0.7	56	0.95	13.937000000000001
2015 Porsche GT3	13.827	93.0	6163	0.71	0.26	70	-2.0	13.827
2011 Ford Mustang GT500	14.035	87.0	2846	0.81	-0.71	47	-2.0	14.035
2015 Porsche GT3	13.827	93.0	6542	0.49	-0.18	76	-2.0	13.827
2014 Chevrolet Corvette	14.624	105.0	4425	0.32	0.05	100	-0.17	14.624
2015 Porsche GT3	13.827	93.0	7763	0.24	-0.18	91	-2.0	13.827

Imputer fills in blanks for medium, or specific value.
| inputlookup track_day_missing.csv
| fit **RobustScaler** * [Link]()

vehicleType	batteryVoltage	engineCoolantTemperature	engineSpeed	lateralGForce	longitudeGForce	speed	verticalGForce	RS_batteryVoltage
2015 Porsche GT3		93.0	6060	1.11	0.5	69	-2.0	
2013 Audi RS5	13.937000000000001	94.0	4957	0.56	0.7	56	0.95	-0.1406926406926397
2015 Porsche GT3	13.827	93.0	6163	0.71	0.26	70	-2.0	-0.37878787878788056
2011 Ford Mustang GT500	14.035	87.0	2846	0.81	-0.71	47	-2.0	0.07142857142857033
2015 Porsche GT3	13.827	93.0	6542	0.49	-0.18	76	-2.0	-0.37878787878788056

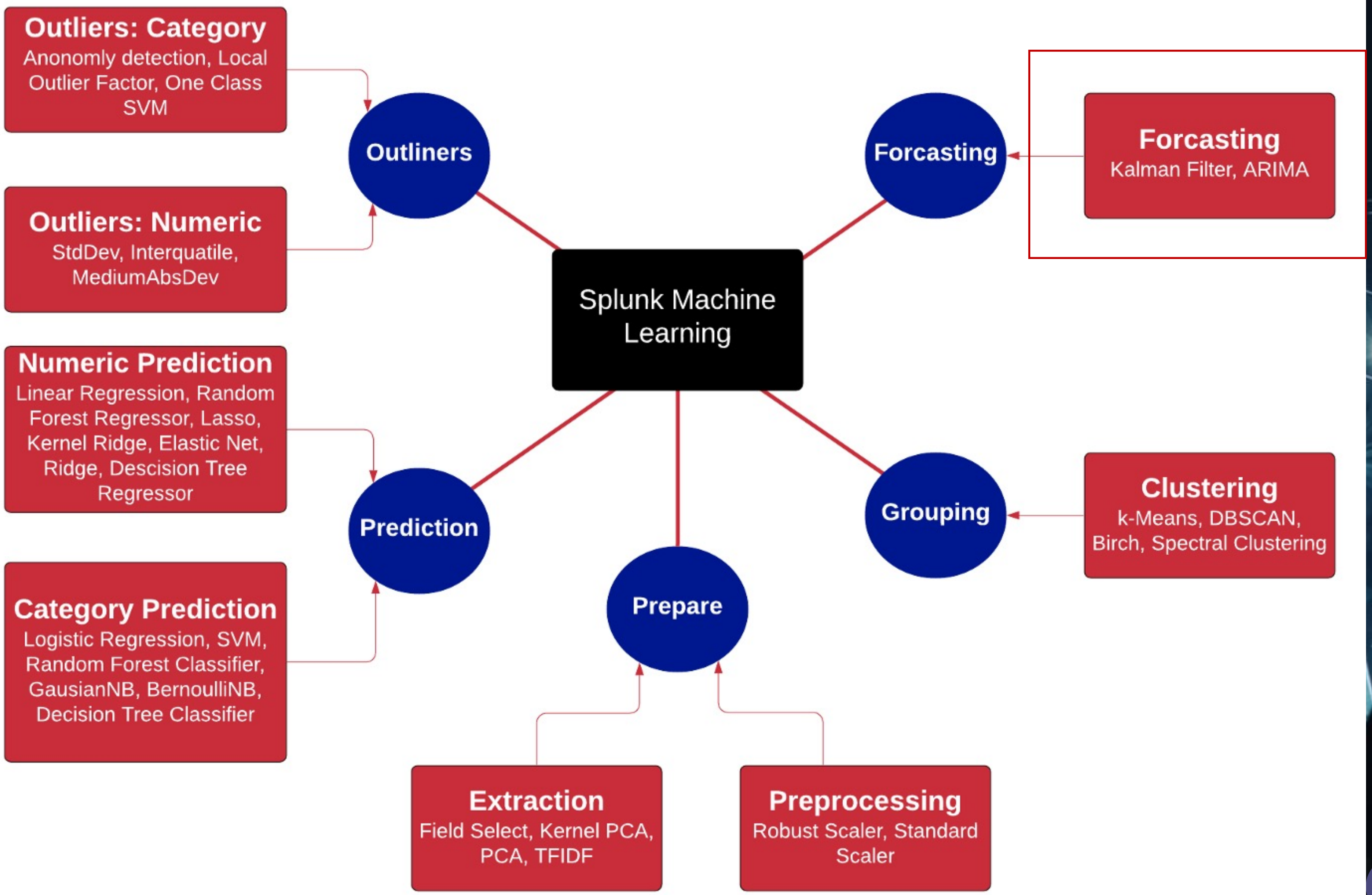
RobustScaler scales to median and interquartile range to 0 and 1. Avoids dominance of fields with large values.

| inputlookup track_day_missing.csv
| fit **StandardScaler** * [Link]()

vehicleType	batteryVoltage	engineCoolantTemperature	engineSpeed	lateralGForce	longitudeGForce	speed	verticalGForce	SS_batteryVoltage	SS_engineCoolantTemperatu
2015 Porsche GT3		93.0	6060	1.11	0.5	69	-2.0		
2013 Audi RS5	13.937000000000001	94.0	4957	0.56	0.7	56	0.95	-0.4981803519277019	0.324647341076498
2015 Porsche GT3	13.827	93.0	6163	0.71	0.26	70	-2.0	-0.803465525894621	0.27818024136906
2011 Ford Mustang GT500	14.035	87.0	2846	0.81	-0.71	47	-2.0	-0.22619901512081622	-0.00062235687556050
2015 Porsche	13.827	93.0	6542	0.49	-0.18	76	-2.0	-0.803465525894621	0.27818024136906

StandardScalar scales to median and interquartile range to 0 and 1. Avoids dominance of fields with large values.

Preprocessing



Forecasting

| inputlookup app_usage.csv
| fit **StateSpaceForecast** CRM ERP Expenses holdback=12 into SF [Link](#).

StateSpaceForecast based on Kalman filters.

| inputlookup logins.csv
| fit **ARIMA** _time logins holdback=0 conf_interval=95 order=0-0-0 forecast_k=5 as AR [link](#).

RemoteAccess	Webmail	_time	lower95(predicted(CRM))	lower95(predicted(ERP))	lower95(predicted(Expenses))	predicted(CRM)	predicted(ERP)
897	760	2015-07-28				787.6566719248075	245.06352282287813
938	904	2015-07-27				760.523128268509	243.80372342031262
859	736	2015-07-21				756.5699826494289	240.70924134242486
853	636	2015-08-04				748.0078713245281	231.7229700844895
828	804	2015-07-29				744.0975823377198	241.92535659112744
902	725	2015-08-11				736.8298191126142	226.07432574518296
953	680	2015-08-19				733.5698420631398	219.37842803297875

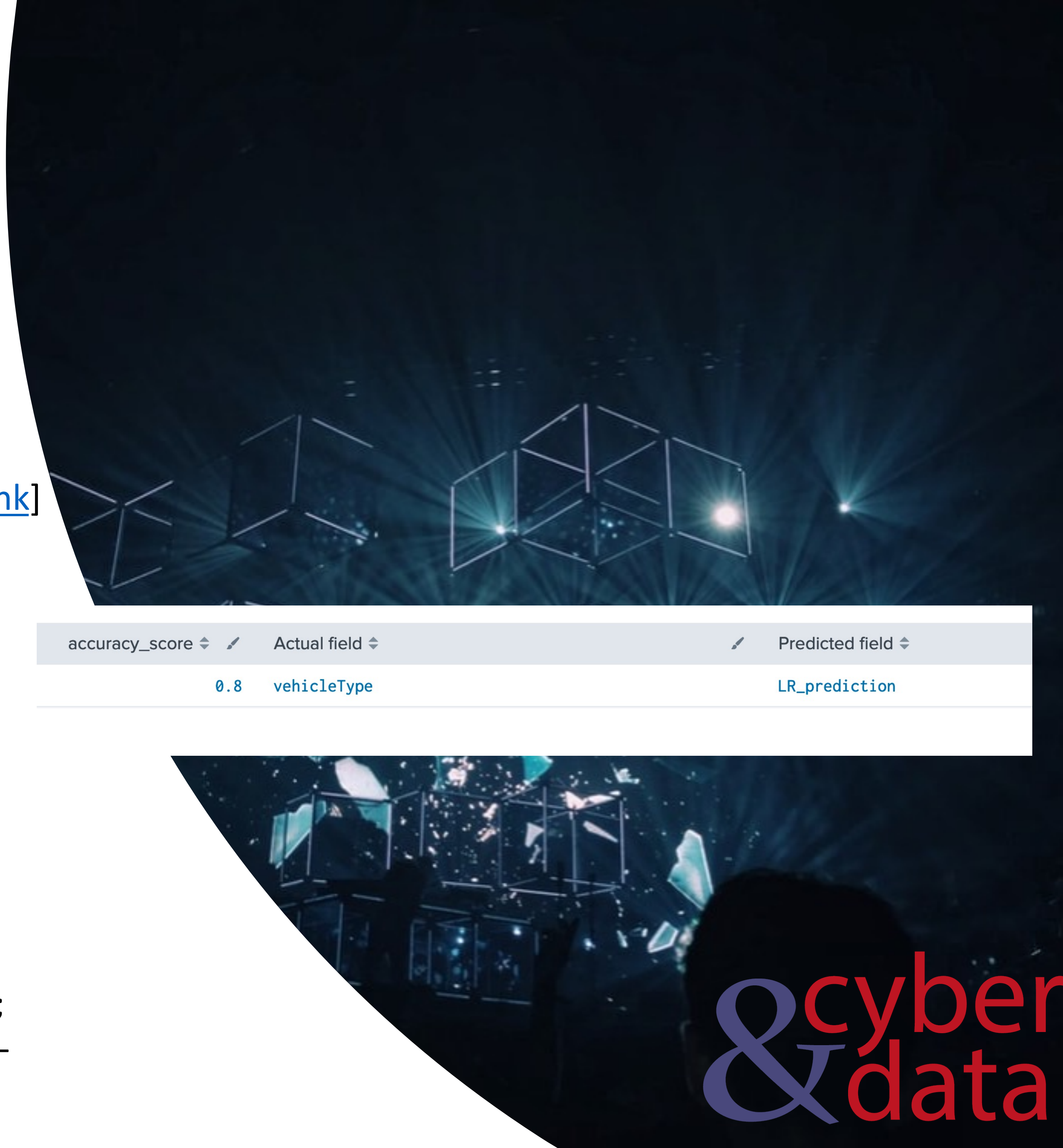
cyber & data

Scoring

- **score. Show scores for tests.**

```
| inputlookup track_day.csv
| head 100
| fit LogisticRegression vehicleType from batteryVoltage engineCoolantTemperature engineSpeed into LR_model
| apply LR_model as LR_prediction
| score accuracy_score vehicleType against LR_prediction
```
[Link]

- **Classification**: Accuracy, Confusion matrix, F1-score, Precision, Precision-Recall-F1-Support, Recall, ROC-AUC-score, ROC-curve.
- **Clustering**: Silhouette score.
- **Pairwise distances scoring:** Pairwise distances score.
- **Regression scoring**: Explained variance score; Mean absolute error score; Mean squared error; and R2 score.
- **Statistical functions** (statsfunctions): Describe; Moment, Pearson; Spearman; Tmean; Trim, and Tvar.
- **Statistical testing** (statstest): Analysis of Variance (Anova); Augmented Dickey-Fuller (Adfuller); Energy distance; One-way ANOVA; and T-test (1 sample).



accuracy_score	Actual field	Predicted field
0.8	vehicleType	LR_prediction

“From bits to information”

Introduction to Splunk and Machine Learning – Part 1