MONASH University

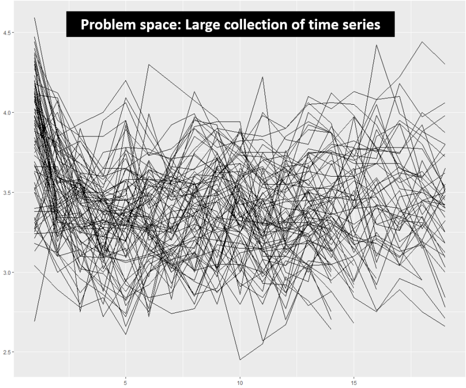
MONASH
BUSINESS
SCHOOL

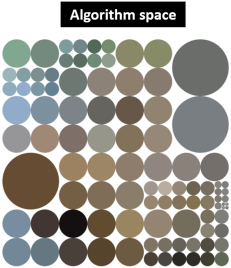
Feature-based Time Series Forecasting

Thiyanga Talagala,
Rob J Hyndman, George Athanasopoulos
Feng Li, Yanfei Kang

11 July 2019

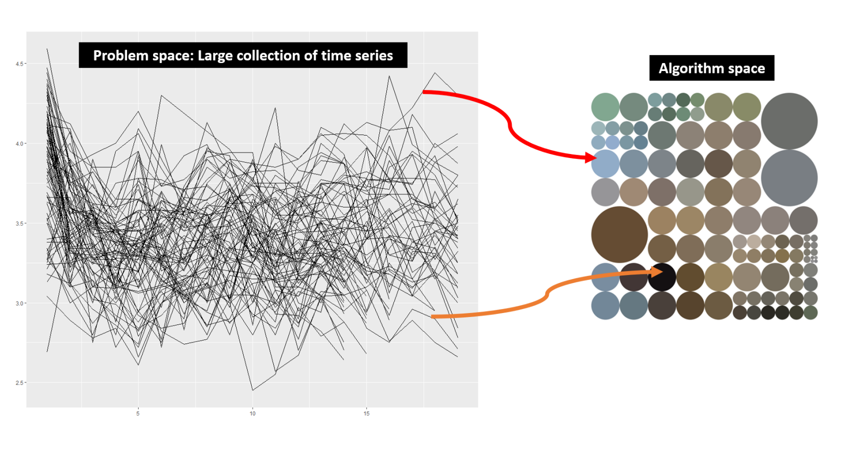
Big picture





- What algorithm is likely to perform best?

Big picture



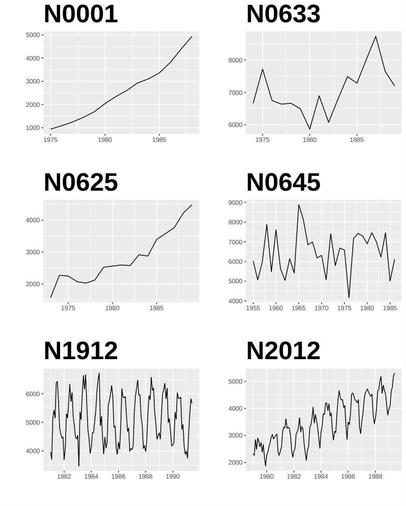
- What algorithm is likely to perform best?
- Algorithm selection problem, John Rice (1976)

Time series features

- Transform a given time series $y = \{y_1, y_2, \cdots, y_n\}$ to a feature vector $F = (f_1(y), f_2(y), \cdots, f_p(y))'\}$.

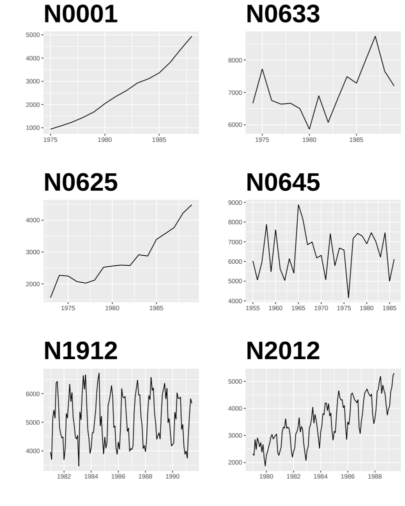
Time series features

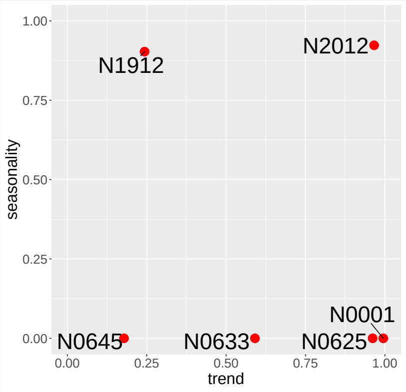
- Transform a given time series $y = \{y_1, y_2, \cdots, y_n\}$ to a feature vector $F = (f_1(y), f_2(y), \cdots, f_p(y))'\}$.



Time series features

- Transform a given time series $y = \{y_1, y_2, \cdots, y_n\}$ to a feature vector $F = (f_1(y), f_2(y), \cdots, f_p(y))'\}$.

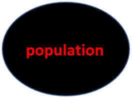




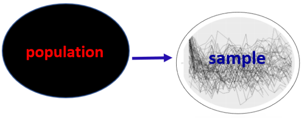
More features

- length
- strength of seasonality
- strength of trend
- linearity
- curvature
- spikiness
- stability
- lumpiness
- spectral entropy
- Hurst exponent
- nonlinearity
- unit root test statistics
- parameter estimates of Holt's linear trend method
- parameter estimates of Holt-Winters' additive method
- ACF and PACF based features - calculated on raw, differenced, seasonally-differenced series and remainder series.

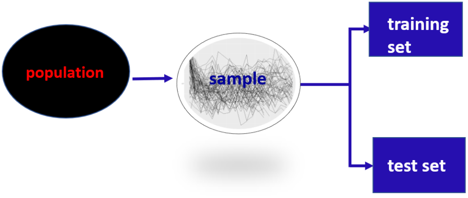
Algorithm selection framework



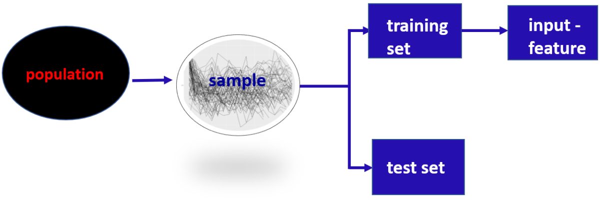
Algorithm selection framework



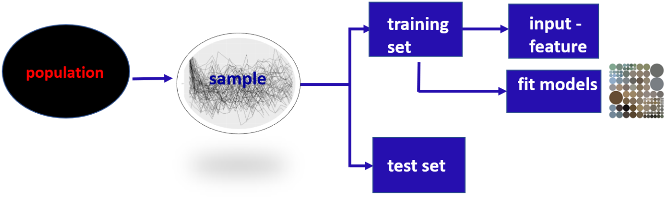
Algorithm selection framework



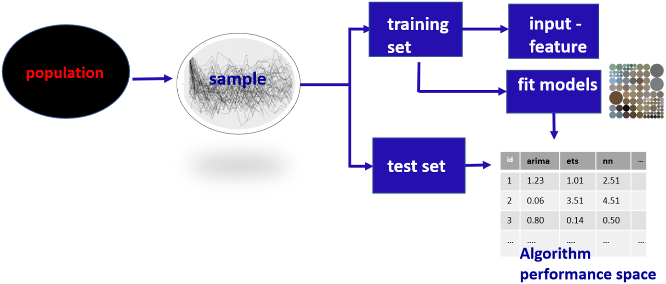
Algorithm selection framework



Algorithm selection framework

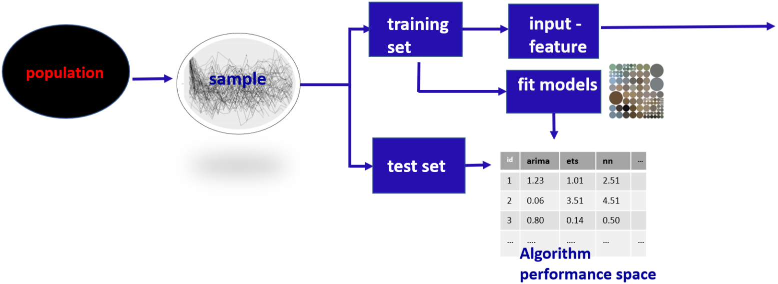


Algorithm selection framework



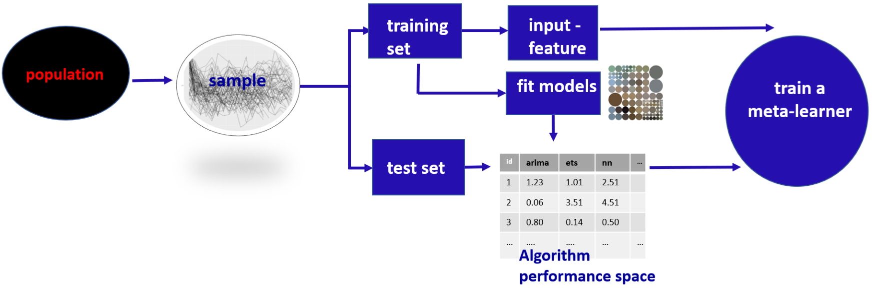
id	arima	ets	nn	...
1	1.23	1.01	2.51	
2	0.06	3.51	4.51	
3	0.80	0.14	0.50	
...	....	....	...	...

Algorithm selection framework

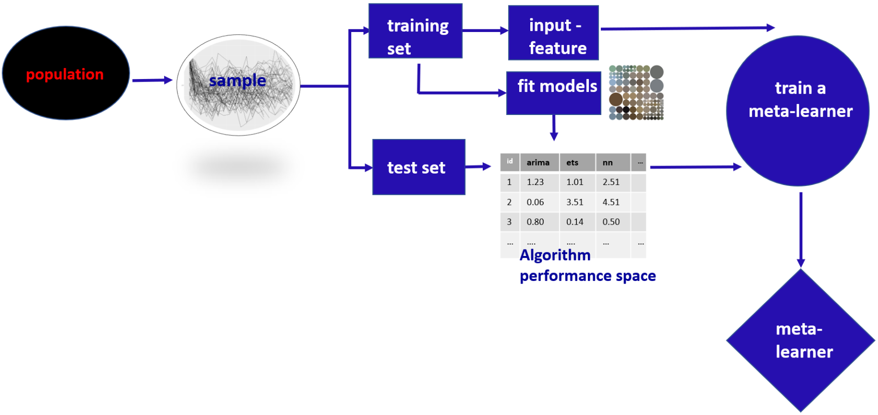


id	arima	ets	nn	...
1	1.23	1.01	2.51	
2	0.06	3.51	4.51	
3	0.80	0.14	0.50	
...	....	....	...	...

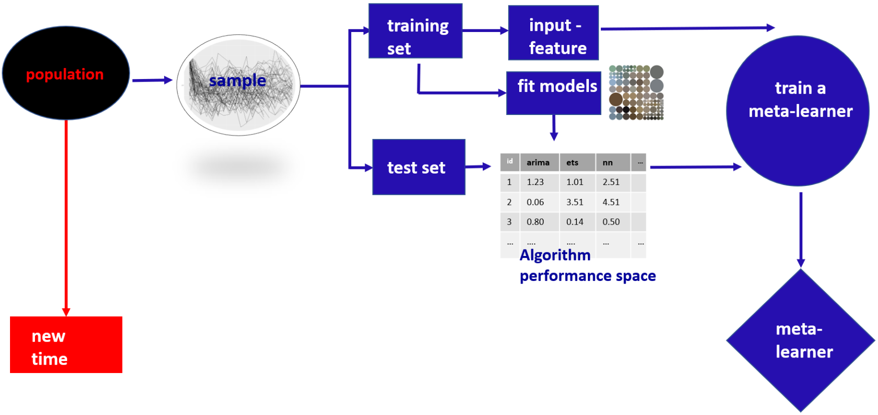
Algorithm selection framework



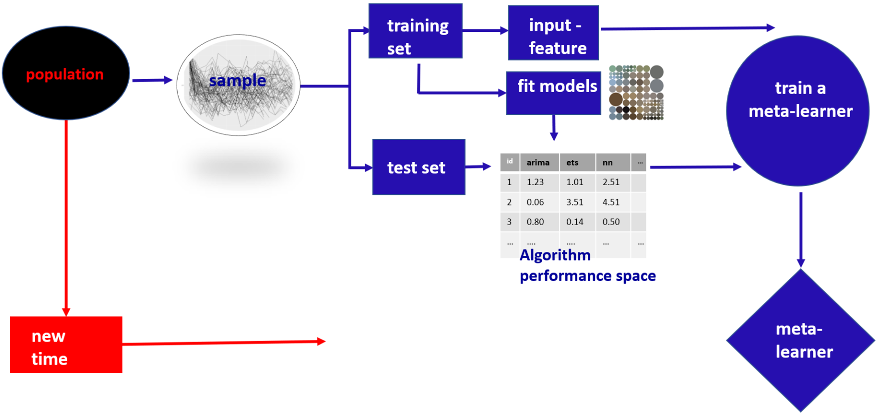
Algorithm selection framework



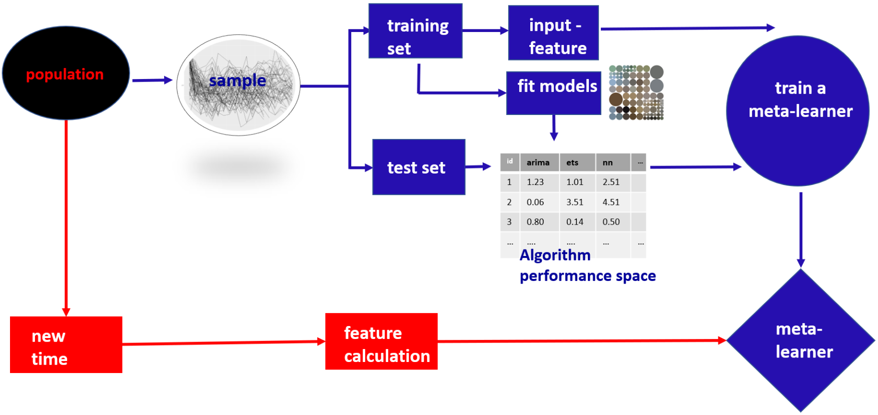
Algorithm selection framework



Algorithm selection framework

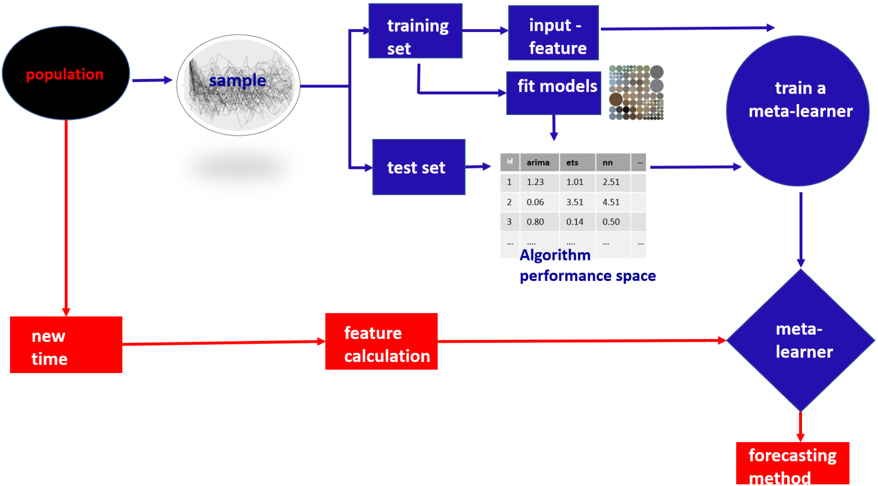


Algorithm selection framework



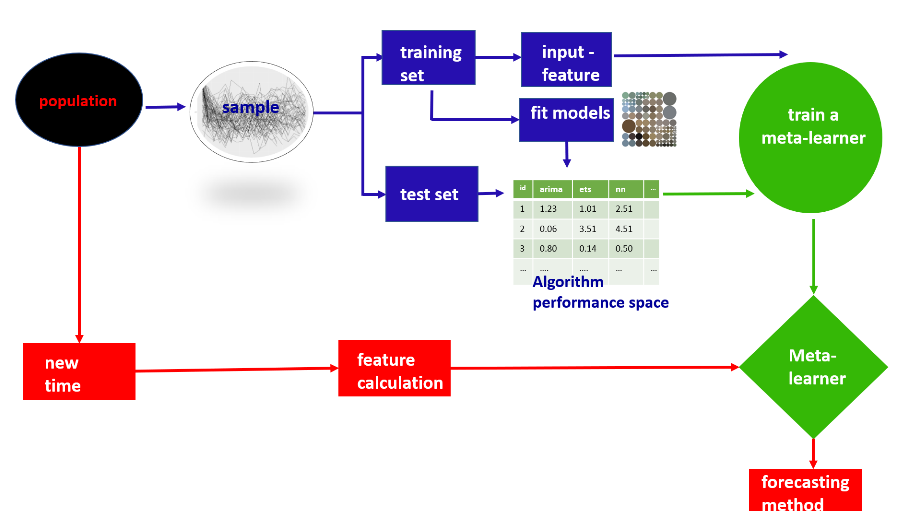
id	arima	ets	nn	...
1	1.23	1.01	2.51	
2	0.06	3.51	4.51	
3	0.80	0.14	0.50	
...	....	....	...	...

Algorithm selection framework

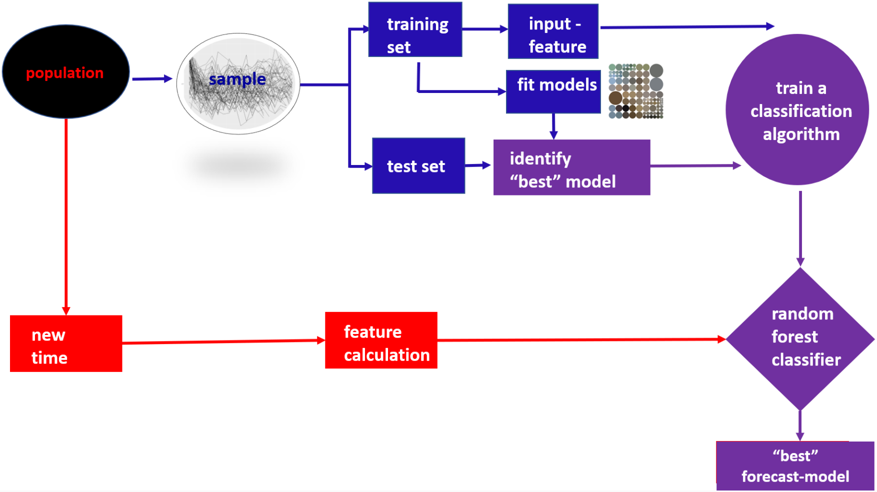


id	arima	ets	nn	...
1	1.23	1.01	2.51	
2	0.06	3.51	4.51	
3	0.80	0.14	0.50	
...	....	....	...	...

Algorithm selection framework

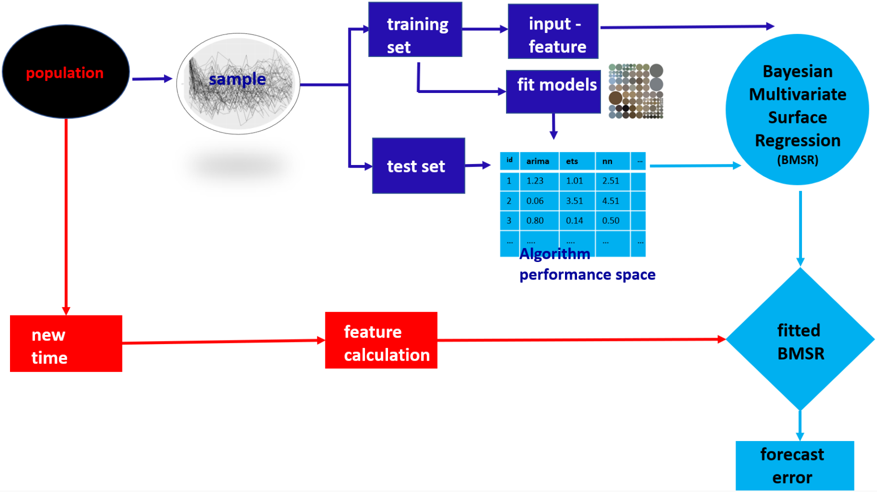


FFORMS: Feature-based FORecast Model Selection



- three algorithms: FFORMS, FFORMPP

FFORMPP: Feature-based FORecast Model Performance Prediction



- three algorithms: FFORMS, FFORMPP

seer R package

Installation

```r
devtools::install_github("thiyangt/seer")
library(seer)
```

seer R package

Installation

```r
devtools::install_github("thiyangt/seer")
library(seer)
```

Example datasets

observed time series - M1 yearly series (181)

```r
library(Mcomp)
yearlym1 <- subset(M1, "yearly")
```

seer R package

Installation

```
devtools::install_github("thiyangt/seer")
library(seer)
```

Example datasets

observed time series - M1 yearly series (181)

```
library(Mcomp)
yearlym1 <- subset(M1, "yearly")
```

Input: features

```r
cal_features(yearlym1[1:3], database="M1",
h=6, highfreq=FALSE)
```

```
# A tibble: 2 x 25
  entropy lumpiness stability hurst trend spikiness linearity curvature
    <dbl>     <dbl>     <dbl> <dbl> <dbl>     <dbl>     <dbl>     <dbl>
1   0.683    0.0400     0.977 0.985 0.985   1.32e-6      4.46     0.705
2   0.711    0.0790     0.894 0.988 0.989   1.54e-6      4.47     0.613
# ... with 17 more variables: e_acf1 <dbl>, y_acf1 <dbl>,
#   diff1y_acf1 <dbl>, diff2y_acf1 <dbl>, y_pacf5 <dbl>,
#   diff1y_pacf5 <dbl>, diff2y_pacf5 <dbl>, nonlinearity <dbl>,
#   lmres_acf1 <dbl>, ur_pp <dbl>, ur_kpss <dbl>, N <int>, y_acf5 <dbl>,
#   diff1y_acf5 <dbl>, diff2y_acf5 <dbl>, alpha <dbl>, beta <dbl>
```

Input: features

```
seer::fcast_accuracy(tslist=yearlym1[1:2],
               models= c("arima","ets","rw","rwd", "theta", "nn"),
               database ="M1", cal_MASE, h=6,
               length_out = 1,
               fcast_save = TRUE)
```

```
$accuracy
         arima       ets       rw       rwd     theta        nn
YAF2 10.527612 10.319029 13.52428 10.527612 12.088375 11.811063
YAF3  5.713867  7.704409  7.78949  5.225965  6.225463  6.700769

$ARIMA
                     YAF2                      YAF3
"ARIMA(0,1,0) with drift" "ARIMA(0,1,1) with drift"

$ETS
        YAF2        YAF3
"ETS(A,A,N)" "ETS(M,A,N)"

$forecasts
$forecasts$arima
         YAF2     YAF3
[1,] 579581.0 390955.9
[2,] 605761.9 407325.1
[3,] 631942.9 423694.4
```

Training set

```r
prepare_trainingset(accuracy_set = accuracy_m1,
feature_set = features_m1)$trainingset
```

```
# A tibble: 2 x 26
  entropy lumpiness stability hurst trend spikiness linearity curvature
    <dbl>     <dbl>     <dbl> <dbl> <dbl>     <dbl>     <dbl>     <dbl>
1   0.683    0.0400     0.977 0.985 0.985   1.32e-6      4.46     0.705
2   0.711    0.0790     0.894 0.988 0.989   1.54e-6      4.47     0.613
# ... with 18 more variables: e_acf1 <dbl>, y_acf1 <dbl>,
#   diff1y_acf1 <dbl>, diff2y_acf1 <dbl>, y_pacf5 <dbl>,
#   diff1y_pacf5 <dbl>, diff2y_pacf5 <dbl>, nonlinearity <dbl>,
#   lmres_acf1 <dbl>, ur_pp <dbl>, ur_kpss <dbl>, N <int>, y_acf5 <dbl>,
#   diff1y_acf5 <dbl>, diff2y_acf5 <dbl>, alpha <dbl>, beta <dbl>,
#   classlabels <chr>
```

FFORMS classifer

```r
rf <- build_rf(training_set = training_set,
                      testset= M3yearly_features,  rf_type="r
                     import=FALSE, mtry = 8)
```

Predictions

```r
table(rf$predictions)
```

```
##         1         2         3         4         5         6         7
## ETS-trend       rwd       rwd       rwd        nn       rwd       rwd
##         8         9        10
##       rwd       rwd       rwd
## 10 Levels: ARIMA ARMA/AR/MA ETS-dampedtrend ... wn
```

FFORMS

```r
rf$randomforest
```

```
## randomForest(formula = classlabels ~ ., data = training_set,
```

Pre-trained classifiers

Load hourly data FFORMS classifier

```
data("hourly_fforms")
```

Forecast hourly time series in the M4-competition

```
fcast.models <- predict(hourly_fforms, features_M4H)
head(fcast.models)
```

```
##     1     2     3     4     5     6
## tbats    nn stlar stlar    nn    nn
## Levels: mstlarima mstlets nn rw rwd snaive stlar tba
```

FFORMPP: