Database Technology Group

TECHNISCHE UNIVERSITÄT DRESDEN

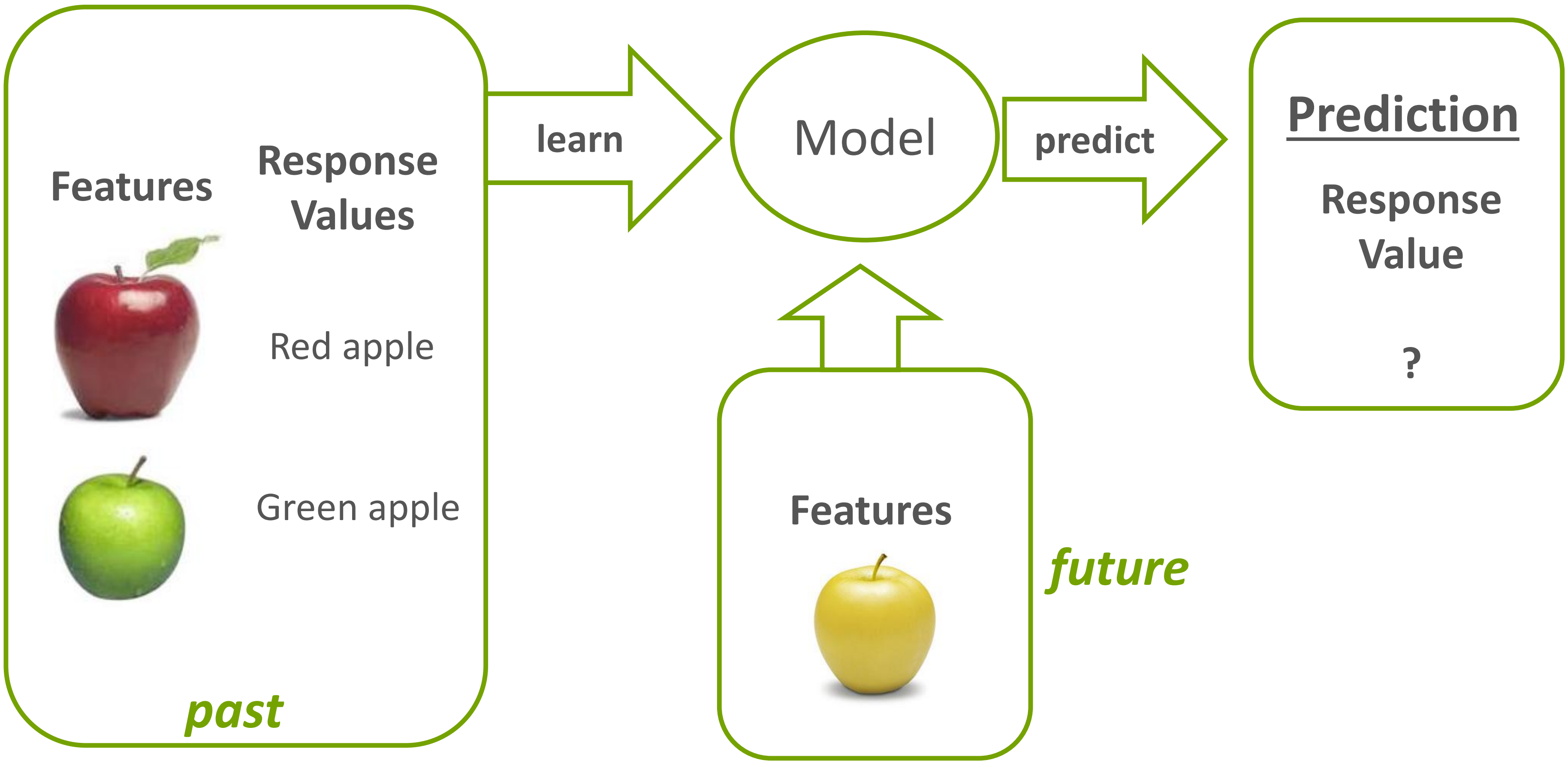
Applicability and Parameterization of Machine Learning Approaches for Time Series Modeling

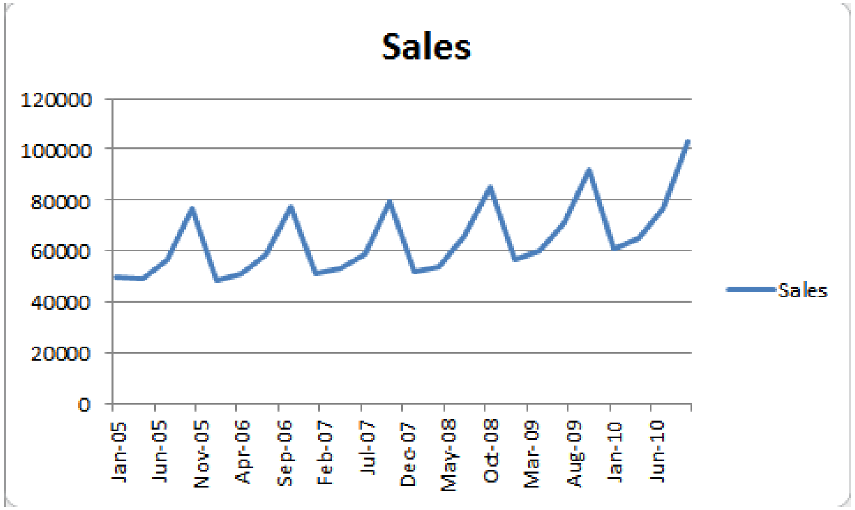
Bachelor thesis – *Onur Ekici*

Montag, 12.05.2014

Machine learning is about predicting the feature based on the past

Hal Daumé III





- *Decision-making and investment planning*
- *Statistical models* → Machine Learning

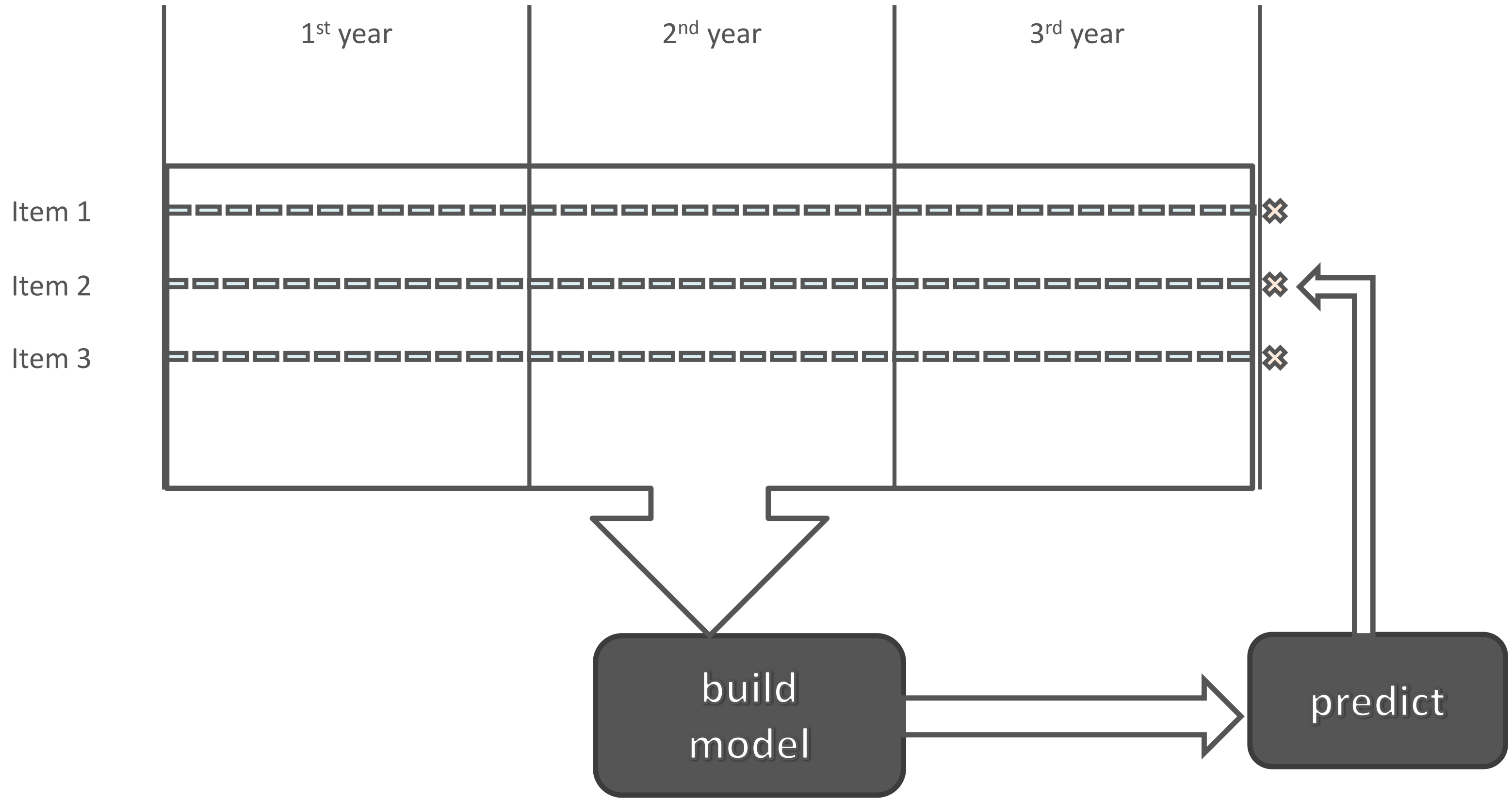
Random Forests

Linear Model

- TIME SERIES MODELING
- MACHINE LEARNING
- BUILD MODEL
- ENSEMBLE MODELS
  - Bias Correction
  - Ensemble Estimations
- CONCLUSION

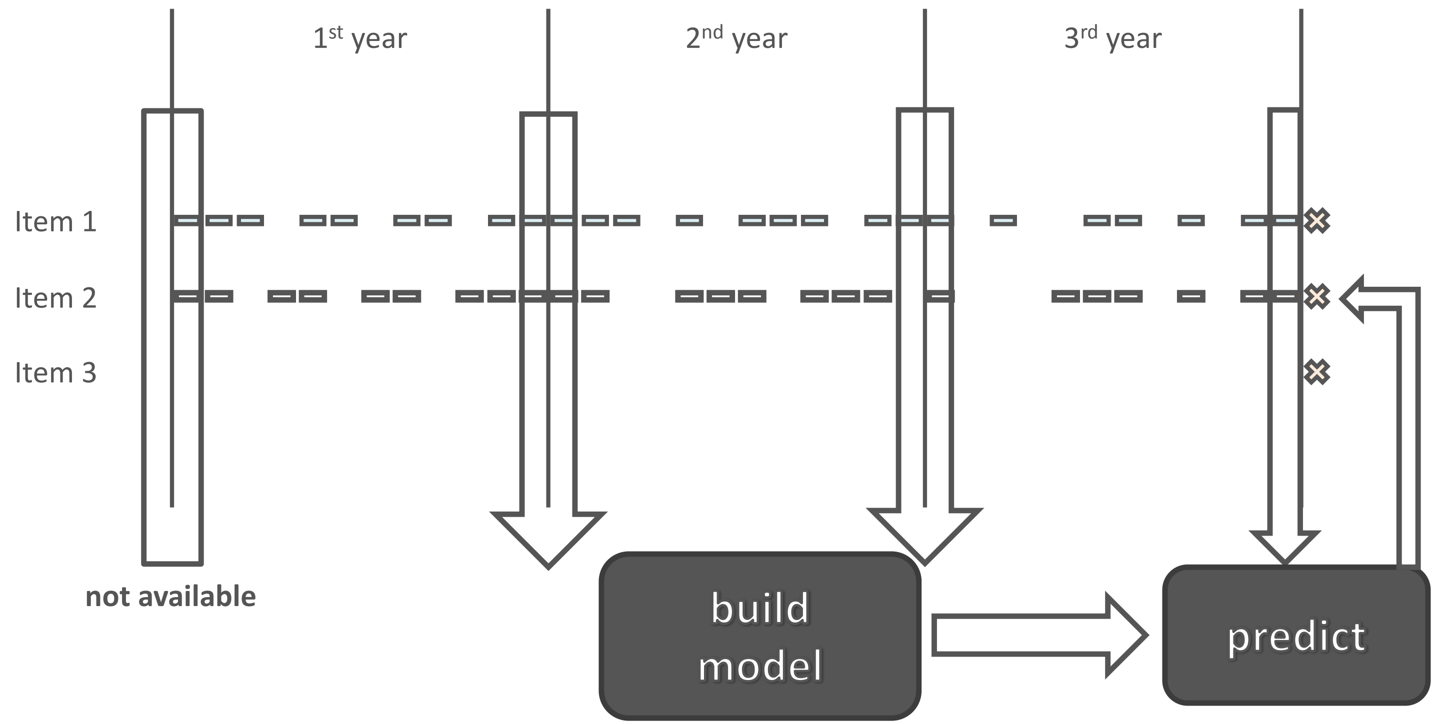
- **TIME SERIES MODELING**
- MACHINE LEARNING
- BUILD MODEL
- ENSEMBLE MODELS
    - Bias Correction
    - Ensemble Estimations
- CONCLUSION

- Generally determined by using statistical models e.g. exponential smoothing, autoregressive model
- Require a long and consistent history



Sparse and too short
for the statistical models

Cross Sectional Forecasting

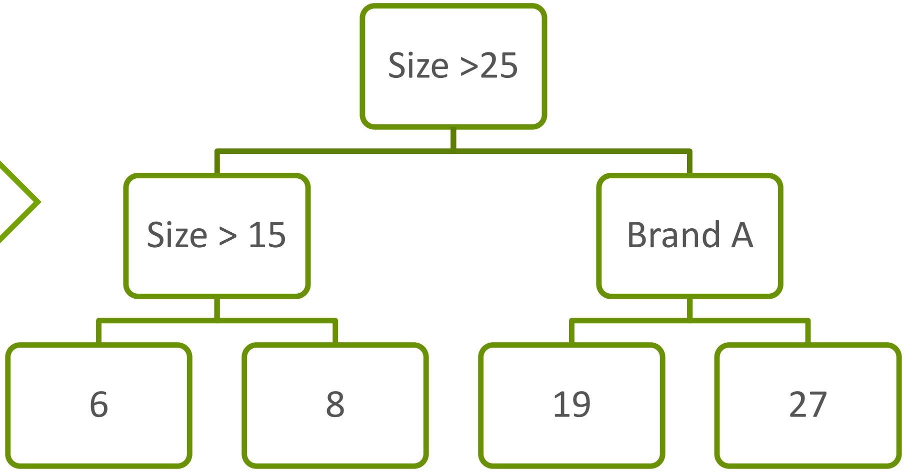


- TIME SERIES MODELING
- **MACHINE LEARNING**
- BUILD MODEL
- ENSEMBLE MODELS
  - Bias Correction
  - Ensemble Estimations
- CONCLUSION

Classification and Regression Tree (CART)

Brand	Size	Price
A	10	6
A	20	8
A	30	19
B	40	27

Train Data



Model → Regression Tree

- *Split the data recursively into two groups to fit*
- *How to split data*
  - Maximum decrease of the impurity of a node → best split
  - Impurity of a node in regression problems:

$$MeanSquaredError = \frac{1}{n}\sum_{i=1}^{n}(Y - Y')^2$$

A simple Example

The ”price” for each instance should be predicted from features ”size” and ”brand”

Brand	Size	Price
A	10	6
A	20	8
A	30	19
B	40	27

$$\frac{60}{4} = 15$$

Before Splitting:

Regression Tree has a single node, which contains all instance

$$MSE = \frac{9^2 + 7^2 + 4^2 + 12^2}{4} = 72.5$$

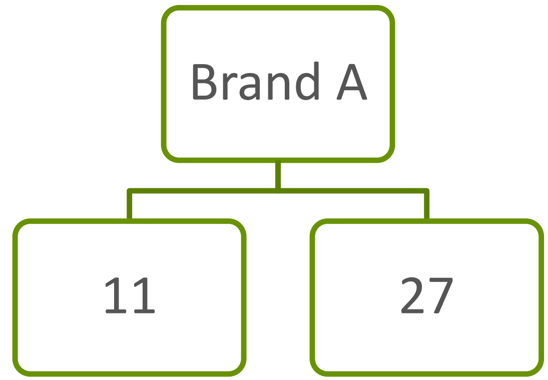
<u>**Now, the aim of CART is to minimize MSE.**</u>

> A simple Example

The "price" for each instance should be predicted from features "size" and "brand"

One possible split :
•A or B

Brand	Size	Price
A	10	6
A	20	8
A	30	19
B	40	27



Finding the best split:
Splitting on brand :

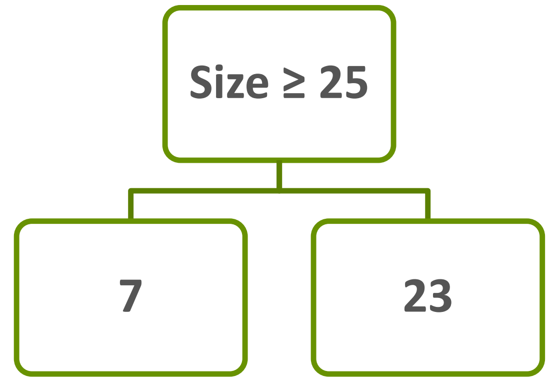
Brand	Size	Price (Y)	Predicted price(Y')	Squared error
A	10	6	11	25
A	20	8	11	9
A	30	19	11	64
B	40	27	27	0
				MSE=98/4=24.5

A simple Example

The "price" for each instance should be predicted from features "size" and "brand"

Three possible split :
- ≥ 15 or not
- ≥ 25 or not
- ≥ 35 or not

Brand	Size	Price
A	10	6
A	20	8
A	30	19
B	40	27



Finding the best split:

The best split for size ≥ 25:

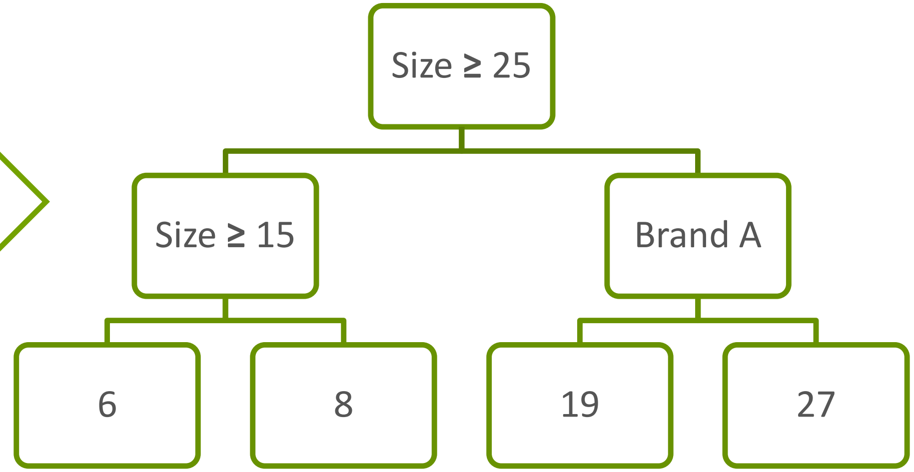
Brand	Size	Price (Y)	Predicted price(Y')	Squared error
A	10	6	7	1
A	20	8	7	1
A	30	19	23	16
B	40	27	23	16
				MSE=34/4=8.5

> A simple Example

- *Same steps again and again.*

Brand	Size	Price
A	10	6
A	20	8
A	30	19
B	40	27

Train Data



Model → Regression Tree

CART
IS
SIMPLE

→

Computationally fast
and
easy interpretable

BUT

- Overfit the training data (when stops fitting ?)
- Small changes lead to big changes in the decision tree

<u>Leo Breiman offers random forest as a solution.</u>

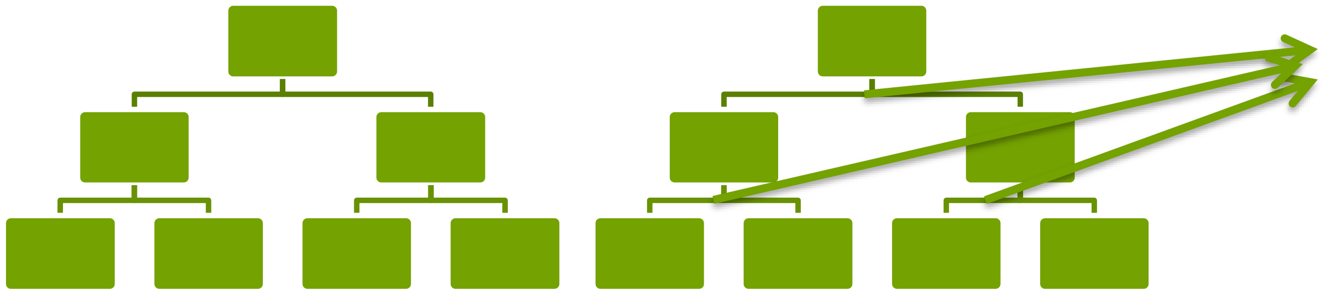
- Random Subspace Method
- Bootstrap Aggregating

CART builds just one tree

Random Forest build lots of different trees with CART Algorithm

Random Subspace Methods

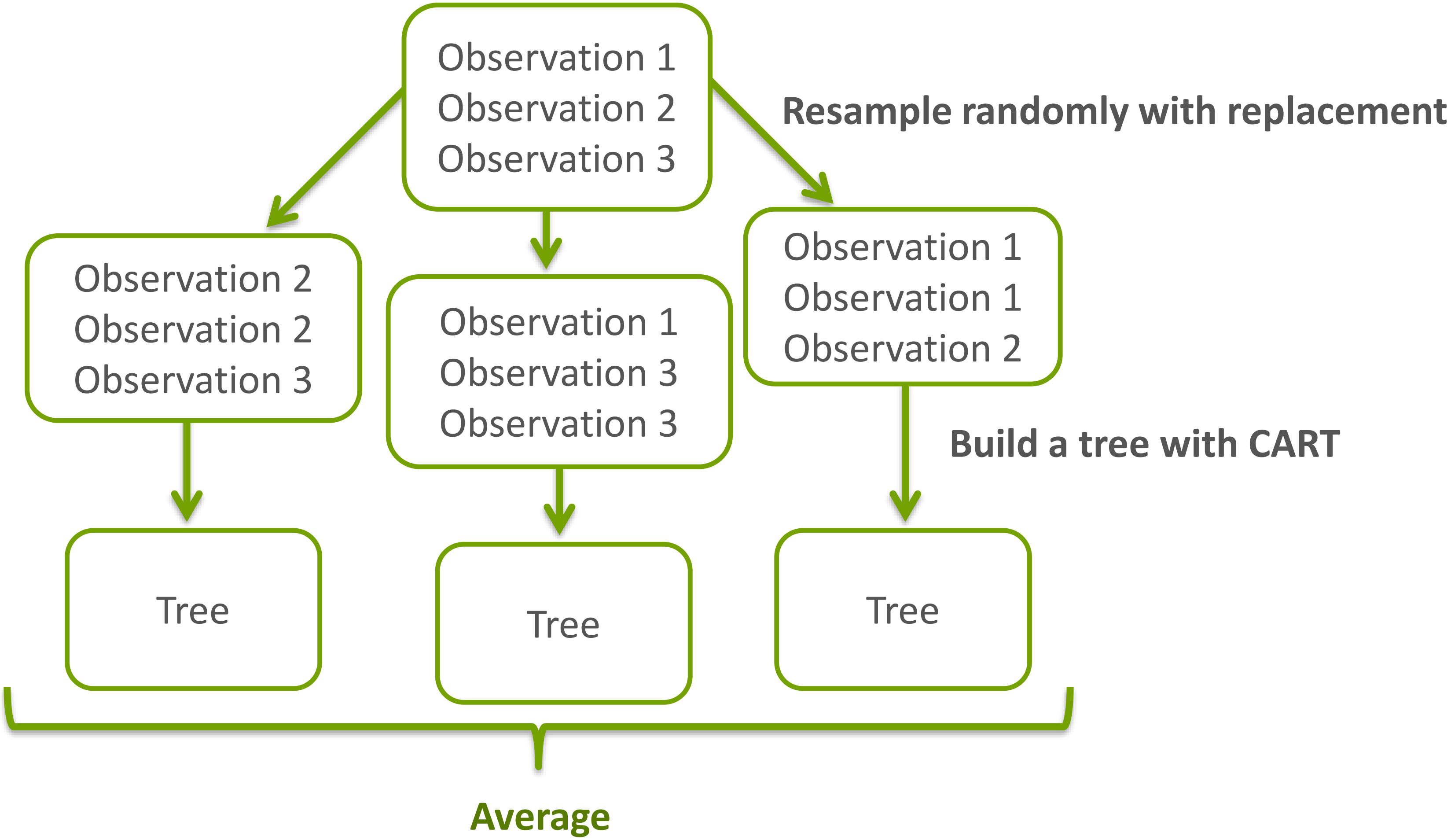
Best split is searched not over all features



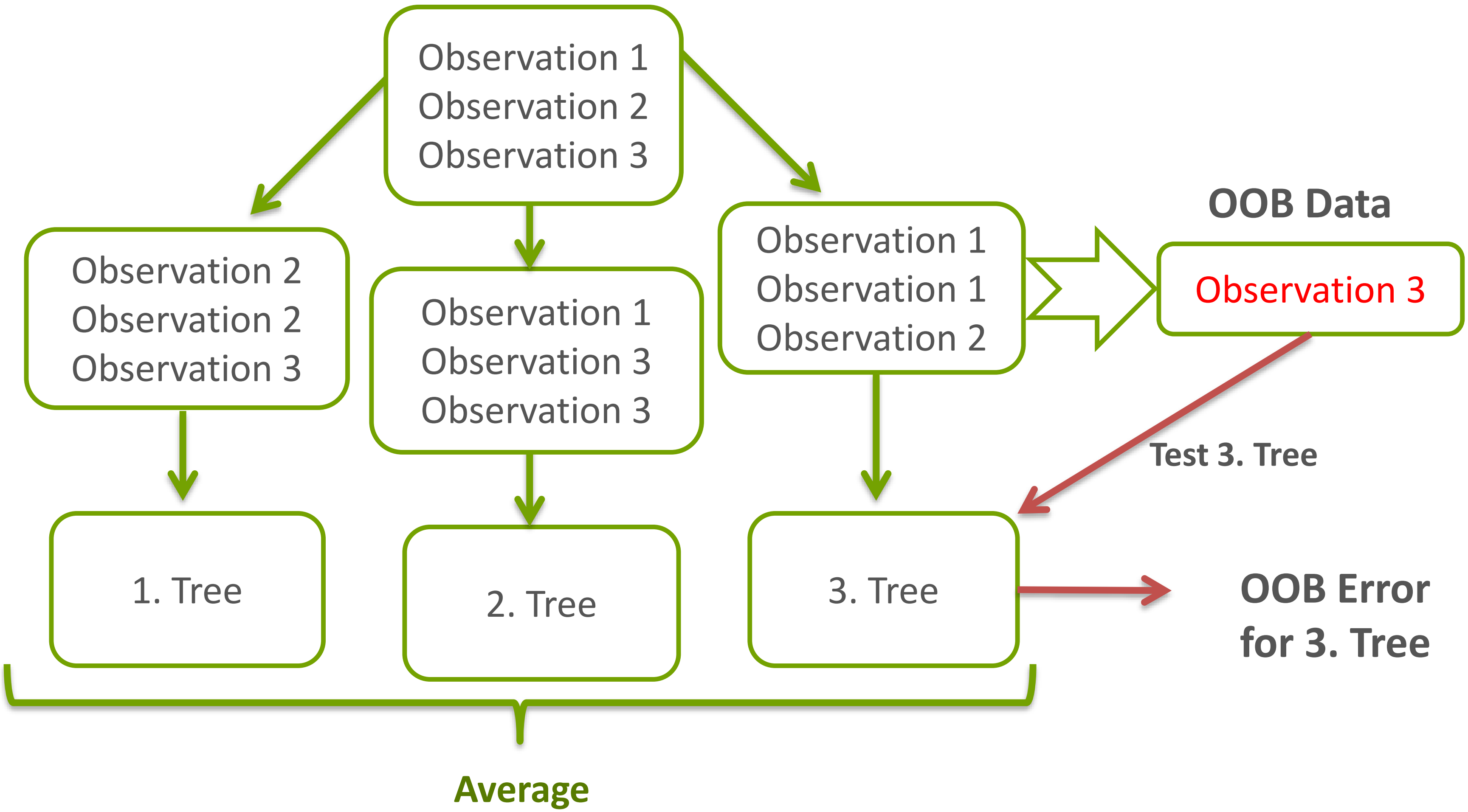
Best split searched from Randomly selected features

> Bootstrap Aggregating

Is a machine learning ensemble method to combine models

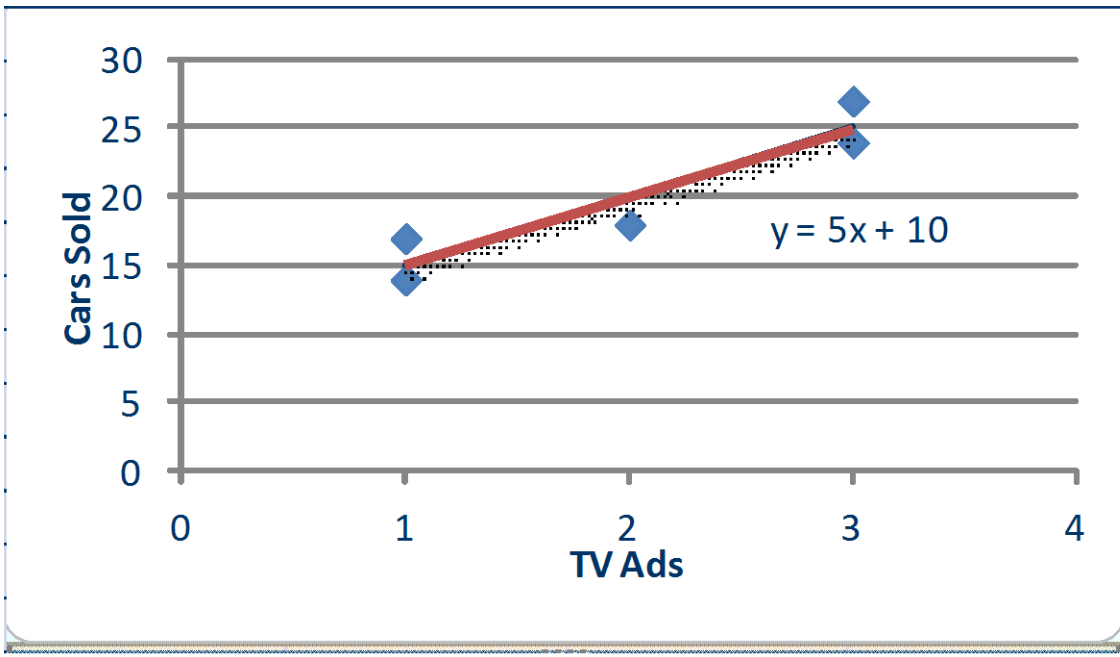


> Out Of Bag Error



> Linear Model

The relationship between a number of independent variables (x1, x2, ...) and a dependent variable y



TV Ads	Cars Sold
1	14
3	24
2	18
1	17
3	27

10 TV advertisement **prediction** 60 cars sold

Example takes from :http://www.coursehero.com/sitemap/schools/3404-Texas-AM-University-Corpus-Christi/courses/235762-ORMS3310/

> Outline

- TIME SERIES MODELING
- MACHINE LEARNING
- **BUILD MODEL**
- ENSEMBLE MODELS
  - Bias Correction
  - Ensemble Estimations
- CONCLUSION

> Representing Task and Data

3 years (36 months) market data is available for this case study

•Building predictive model (first 13 months)
•Evaluation of the model (remaining 22 months)

Following features avaible for each item:
•Sales units in the previous month
•Stock units in the previous month
•Purchase units in the previous month
•Property of Item 1 ... 6

The aim is to predict:
1. Sales number per month
2. Sales number for each brand per month
3. Sales number for each item per month

Identifying relevant predictor variables, rather than only predicting the response by means of some black-box model, is of interest in many applications.

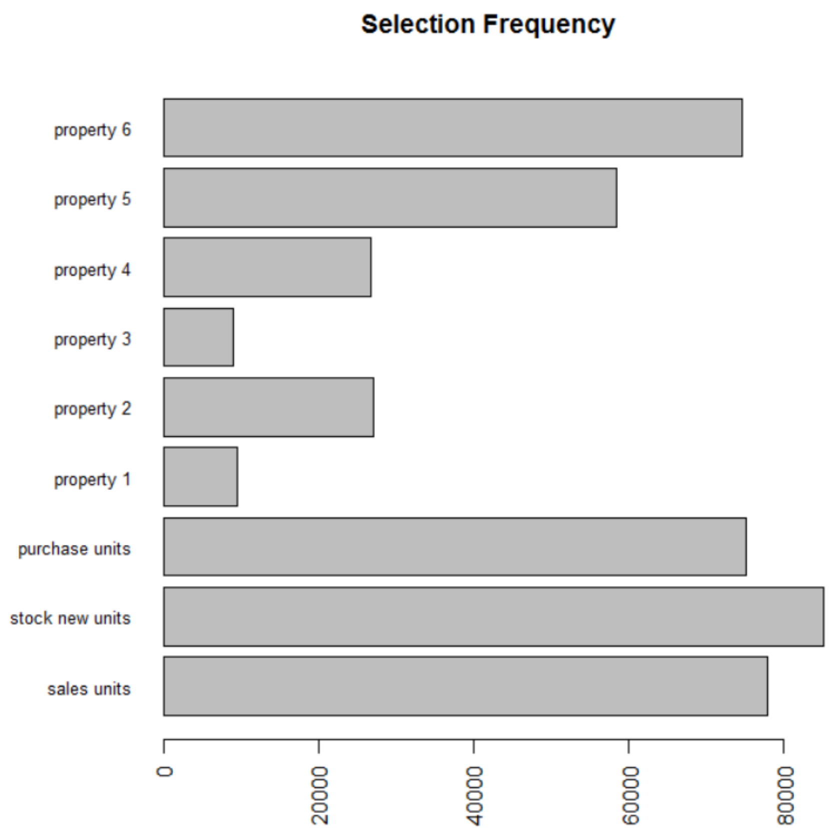
Carolin Strobl 2008

2 possible methods for random forest in regression problems:

1. Selection Frequency
2. Permutation Importance

> Selection Frequency

how offen was each feature used in the individual trees for the division



The relevant Features:

- Property 6
- Property 5
- Purchase units
- Stock new units
- Sales units

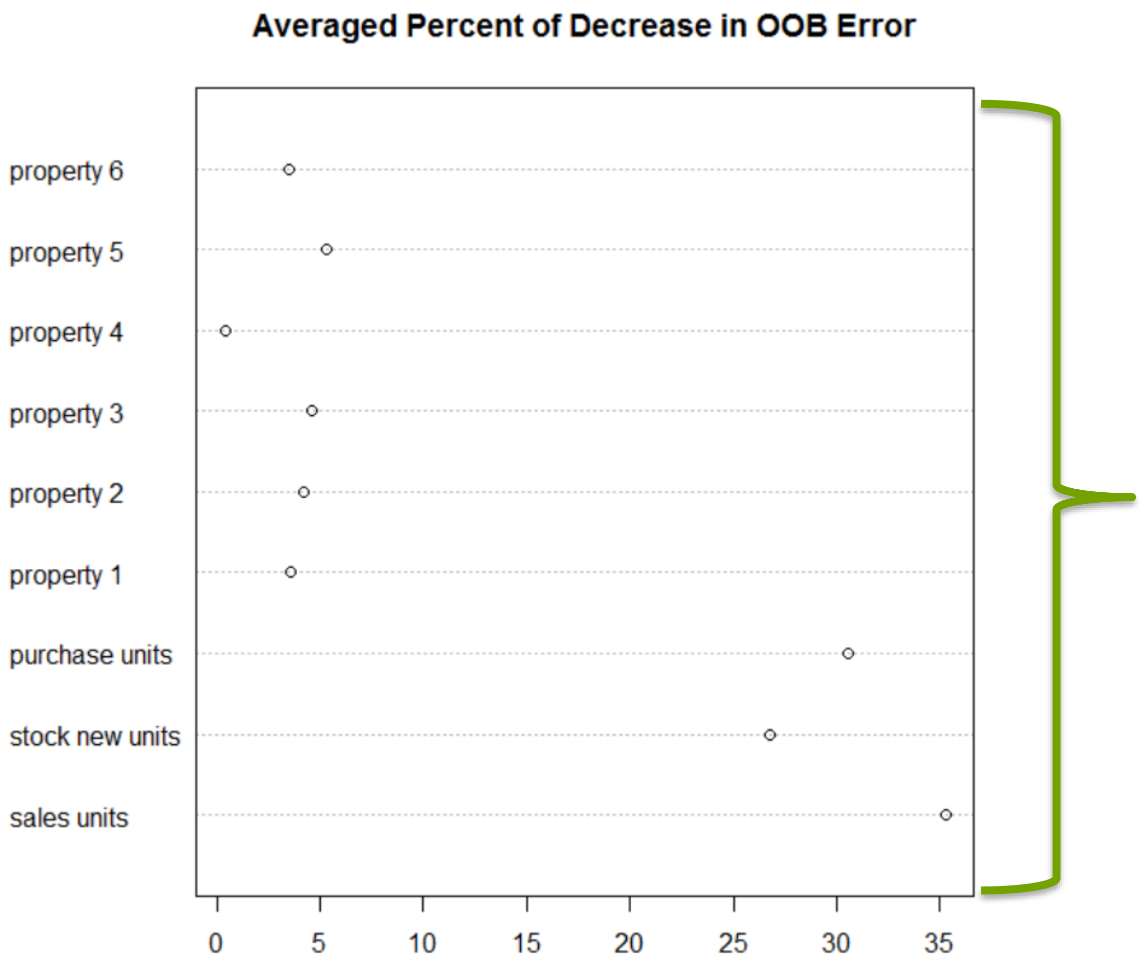
Permutation Importance

The value of the feature is artificially noised and the change of OOB error is measured.

Iterate
for each tree
and
for each feature

1. Calculate OOB error of the tree
2. Permute the feature in the OOB data
3. Calculate again OOB error of the tree
4. Calculate the difference between the first and second OOB error

> Permutation Importance



The relevant Features:

- Purchase units
- Stock new units
- Sales units

> Selecting Features for Random Forest

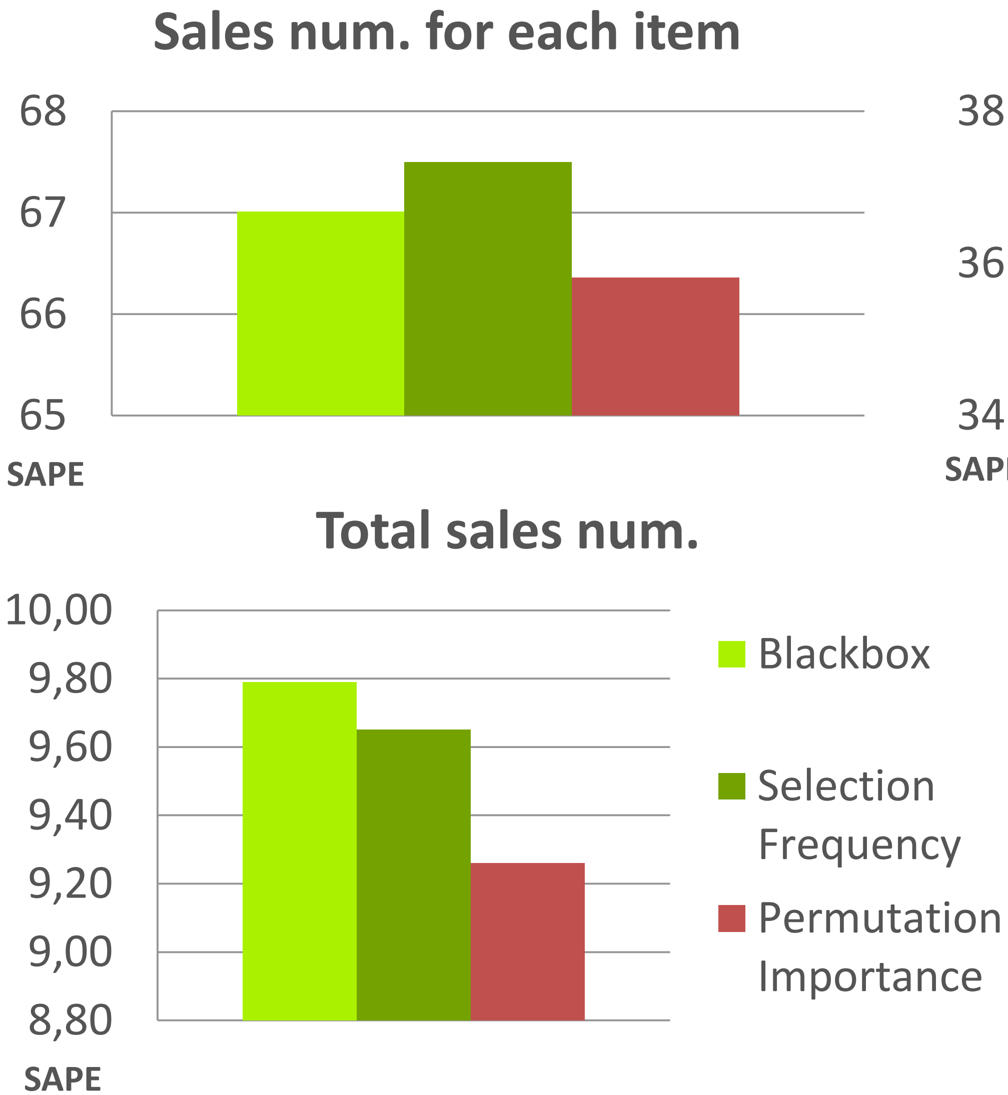
Now there are 3 possible scenarios:

1. *Random Forest with all features (black box model)*
2. *Random Forest with the relevant features through selection frequency*
3. *Random Forest with the relevant features through permutation accuracy*

1. Question is:

Which feature selection method should be used in random forest and is it necessary ?

> Selection of The Relevant Features



- *The permutation importance is more reliable*

- *Choosing relevant features improves the accuracy of random forests.*

- TIME SERIES MODELING
- MACHINE LEARNING
- BUILD MODEL
- **ENSEMBLE MODELS**
  - **Bias Correction**
  - **Ensemble Estimations**
- CONCLUSION

- *The bias is a systematic error in the model estimation.*

Model:

$$y = 5 * x + 10$$

TV Ads	Sold Car
10	65
20	125
30	165

+5

2 methods are introduced to correct bias:

1. *Ensemble of random forest estimation and bias correction with linear model*
2. *Ensemble of random forest estimation and bias correction with random forest*

- *Proposed by Zhang and Lu (2012)*

Linear relationship between the real value and estimated value

$$newEstimation = b_0 + b_1 * Y_{randomForest}$$

Random forest trains → Random forest predicts in OOB Data → Linear model trains in OOB Data

> Bias Correction with Random Forest

- *Proposed by Ruo Xu (2013)*

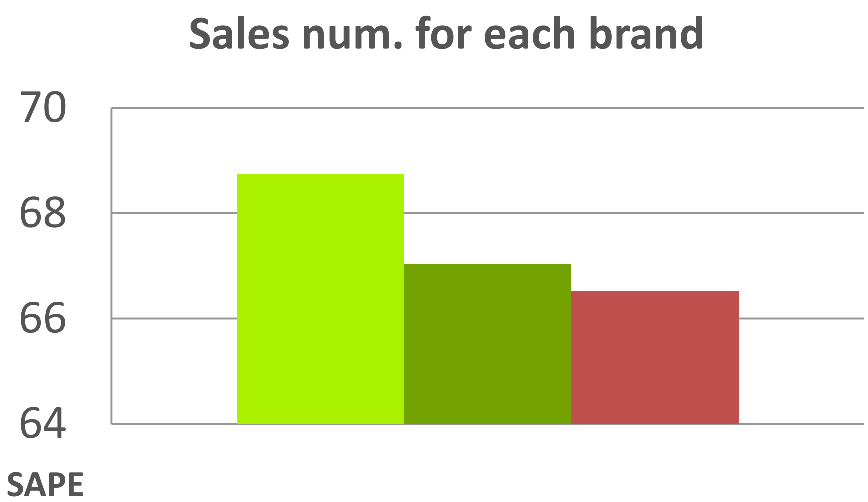
- ***Relationship between the features and bias***
- *Use a second random forest predict bias of the first random forest*

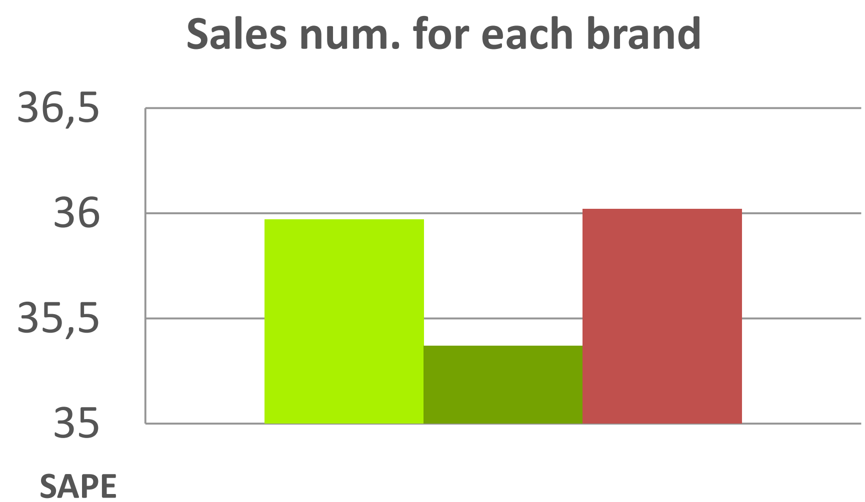
1. Random forest trains → 1. Random forest predicts in OOB Data → Calculates Bias of 1.RF in OOB Data → 2. Random forest trains In OOB Data

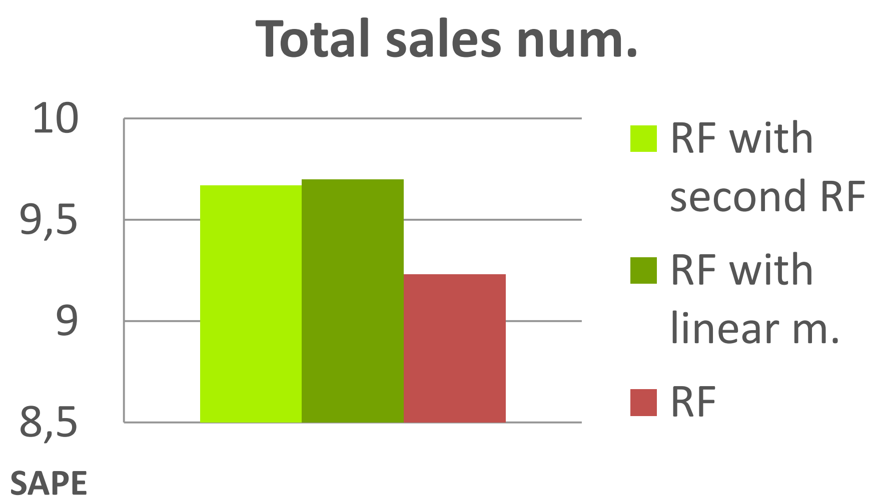
1. *Random forest*
2. *Ensemble of random forest estimation and bias correction with linear model*
3. *Ensemble of random forest estimation and bias correction with random forest*

2. Question is:

How effective are the bias correction methods in time series ?







- *The bias correction methods have not yielded any significant improvement*

> Ensemble of Estimations

Perlich proved that logistic regression and decision trees act as a complement to each other

- *Here the estimations of each model will be equal-weighted combined*

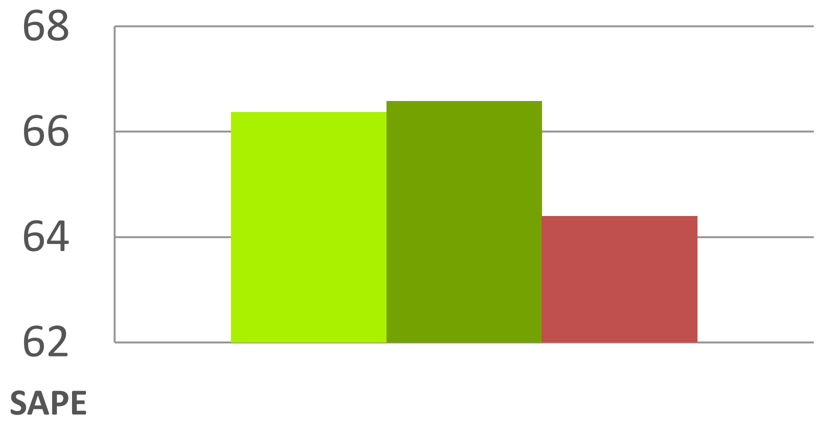
Random forest trains and predicts → Linear model trains and predicts → Average

3. Question :

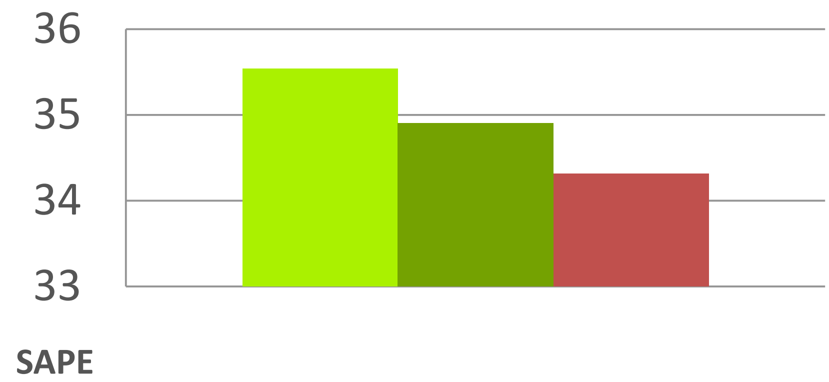
Is it possible to improve accuracy with ensemble of linear model and random forest ?

> Ensemble of Models

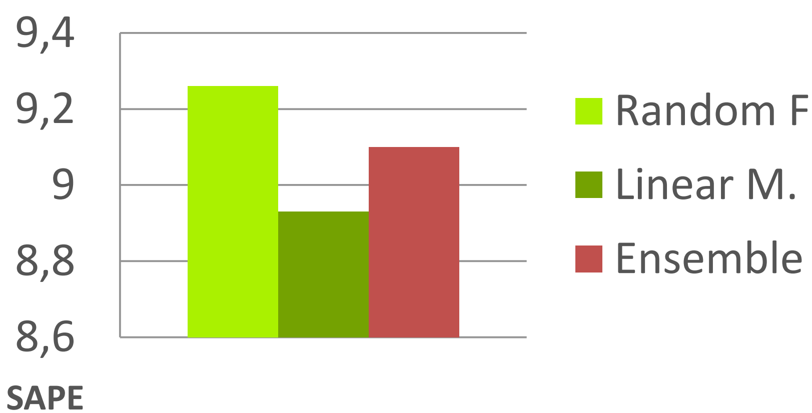
Sales num. for each item



Sales num. for each brand



Total Sales num.



- *The ensemble has a better accuracy than the linear model and the random forest*

- TIME SERIES MODELING
- MACHINE LEARNING
- BUILD MODEL
- ENSEMBLE MODELS
    - Bias Correction
    - Ensemble Estimations
- **CONCLUSION**

> Conclusion

8 different model based on 2 machine learning approaches;

Random Forest and Linear Model

1. *Better to select relevant features instead of some black-box model*
2. *Permutation Importance is more reliable method to select features*
3. *Ensemble of RF and LM has better accuracy than either RF, or LM alone*
4. *Bias correction methods have not resulted any improvement*

Future Works

- *Combine different models: SVM, neural networks etc.*
- *Combine witfh different ratios*

Applicability and Parameterization of Machine Learning Approaches for Time Series Modeling

Bachelor thesis – Onur Ekici

Montag, 12.05.2014

Parameter of Random Forest

There are 3 parameter for random forest:

<u>Number of Trees (ntree) :</u>

the optimization of the performance rather than the optimization of the accuracy.

It should not be set too small,

hence the forest can be stabilized. It should be at least several hundred.

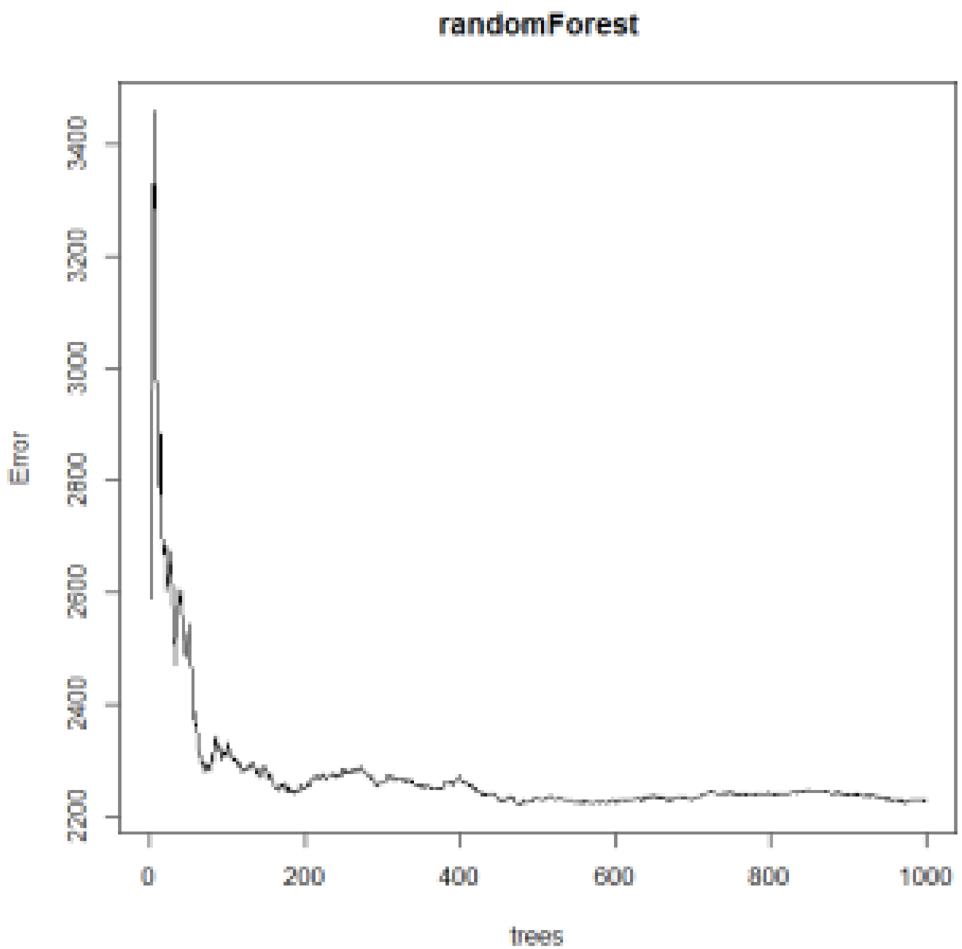
In case it is too large,

the computation takes more time, but the result does not change.

Default value : 500

In this work: 1000

> Parameter of Random Forest



(a) between 1st month - 2nd month

Test was performed in the initial data.

The forest stabilizes after 200 trees in most cases.

Parameter of Random Forest

<u>Node size (nodesize) :</u>

When should stop CART algorithm ?

If the number of instances in the node is less than or equal to the node size, then the algorithm stops splitting and this node is called terminal node.

<u>Important parameter in CART, but has not great effect in random forest.</u>

<u>Default Value : 5</u>

<u>In this work: 5</u>

<u>Number of features sampled (mytr):</u>

It is a key parameter to optimize accuracy of random forest.

Breiman suggested one third of the number of features for regression problems.

The tuneRF function in R Implementation of random forest tunes mytr.