Models with Models

Meta Learners

Agenda

1. Where in the “prediction process” are we talking about
2. Some key terms
3. Sampling and Model Choice *(Example code)*
4. Ensembles
5. Bagging *(Example code)*
6. Boosting *(Example code)*
7. Vertical Stacking *(Example code)*
8. Pipelining

Kicking Off

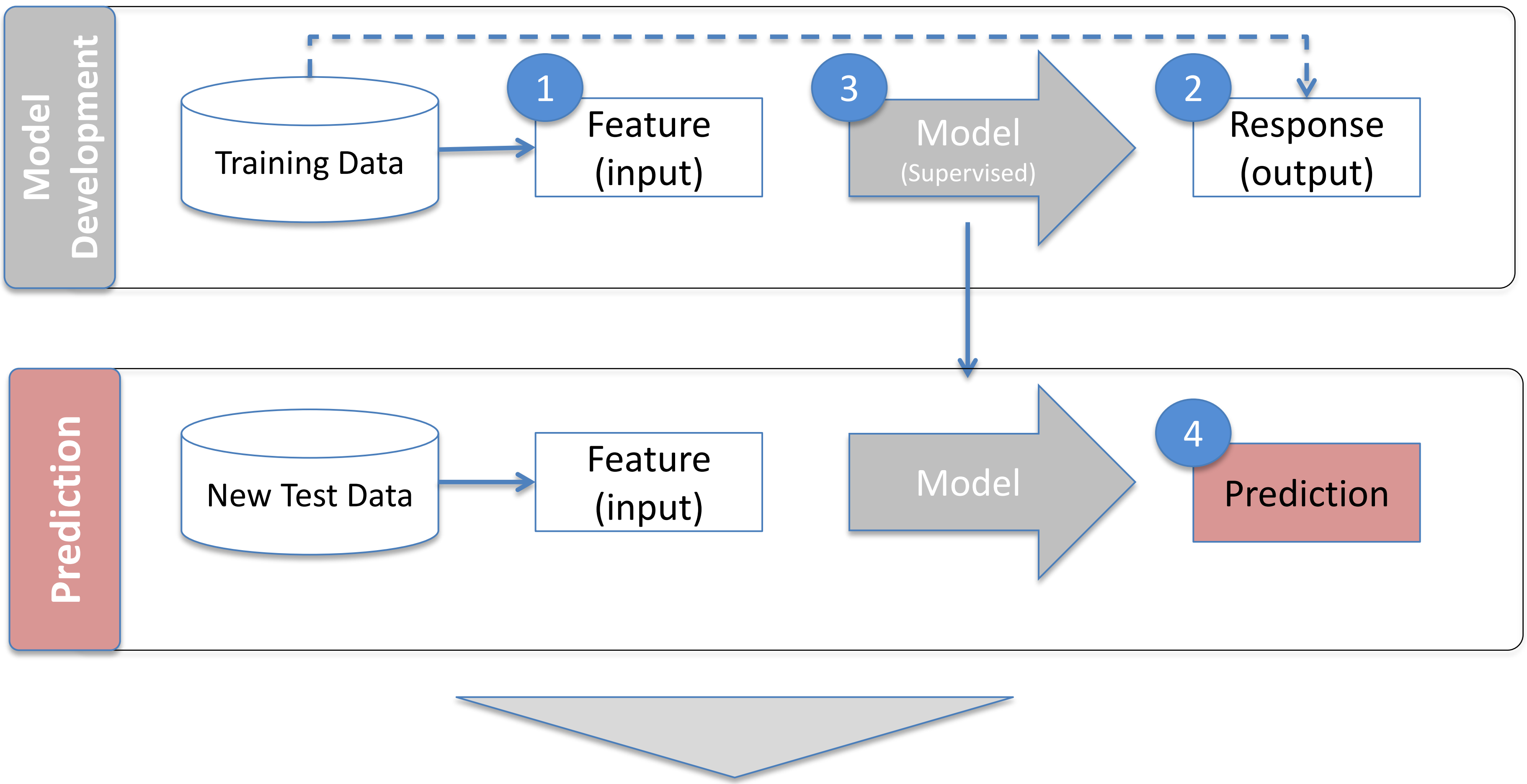
When thinking about the process to develop a predictive model, there are key steps that occur: data acquisition, feature engineering, feature selection, model development (training/tuning/evaluation), and finally prediction on new data. We usually think of the model development step as a single model (i.e. regression, decision tree, SVM, etc.).

1. Business Question / Objective
2. Data collection
3. Feature development /engineering
4. Feature selection
5. Model development
6. Prediction
7. Intervention deployment (w/ DOE)
8. Evaluate
9. Learn and adapt (feedback loop)

My focus will be meta model development (combining and stacking), where the final algorithm may not be a single model, but a series of models that collaboratively drive more accurate predictions.

Sample Code on GitHub: https://github.com/mshump/models_with_models

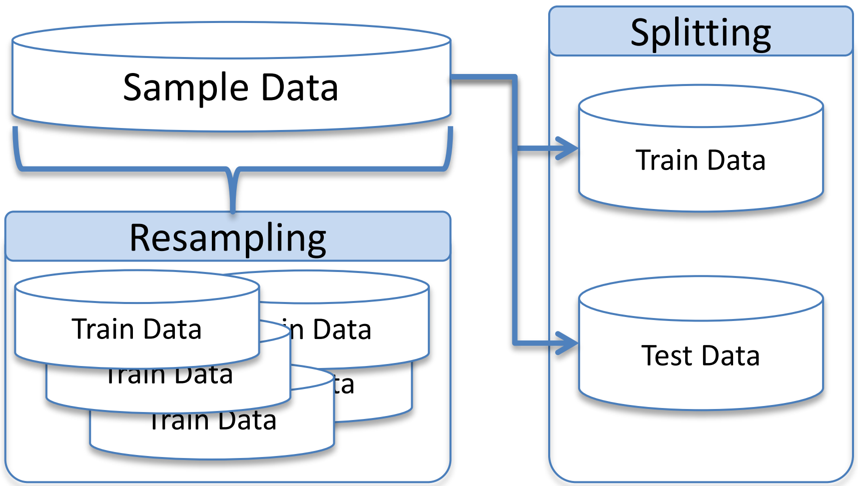
Supervised Learning and Prediction – What is…?



5 Is the Prediction Model Good?

Sampling and Model Choice

Data Sampling



Model Choice

1. Match the Model to the Signal in the noise
2. Tune the model to fit as best it can
3. Evaluate the model's ability to generalize to more data

Some Classification Model Options*

Linear classification

1. Multinomial Logistic regression
2. Linear discriminant

Non-linear classification

3. Mixture Discriminant
4. Regularized Discriminant
5. Flexible Discriminant
6. Neural Network
7. Support Vector Machine
8. K-nearest Neighbors
9. Naïve Bayes

Non-Linear Classification and Decision Trees

10. Classification and Regression Trees (CART)
11. C4.5
12. PART

Bagging and Boosting Models

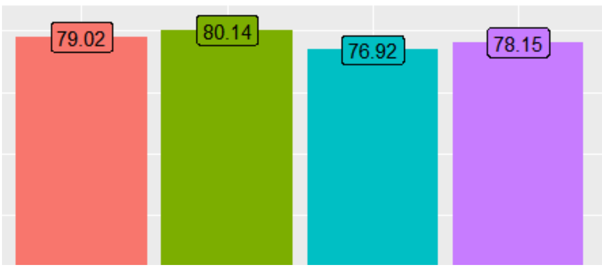
13. Bagging CART
14. Random Forest
15. Gradient Boosted Model
16. Boosted C5.0

Example see code: https://github.com/mshump/models_with_models

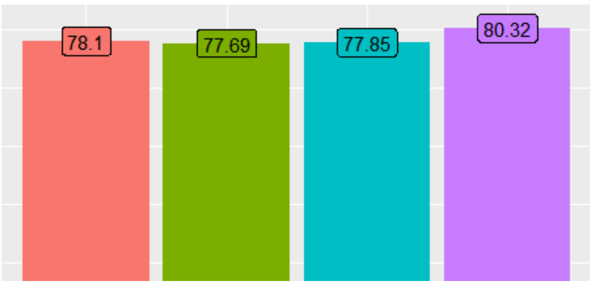
R Sample Code Output (Example 1 & Example 2)

Splitting – Binomial Classification

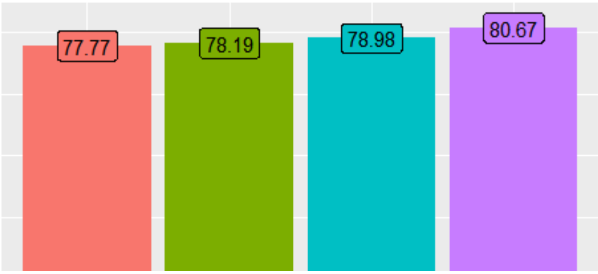
Three different random splits return 3 different prediction performance results



Seed 1: Discriminant > Logistic >>



Seed 2: CART > Logistic >>

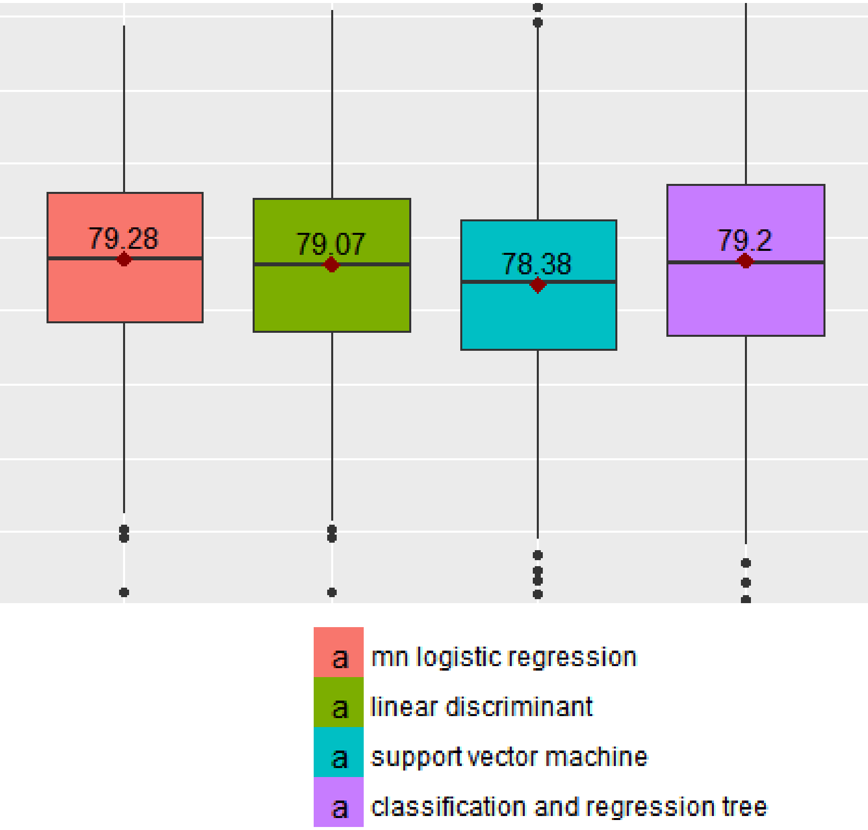


Seed 3: CART > SVM >>

Resampling – Binomial Classification

1000 resamples show close prediction performance amongst 4 models with variation, but suggest different results than from one split of the sample data.

Logistic > CART > Discriminant > SVM



Ensembles – Models as Features

- Ensemble learning involves combining multiple predictions derived by different models, data sets or parameters in order to create a stronger overall prediction.

- For example, the predictions of a random forest, a support vector machine, and a simple linear model may be combined to create a stronger final prediction set.

Some general ensemble methods:

- **Bagging**. Building multiple models (typically of the same type) from different subsamples of the training dataset.

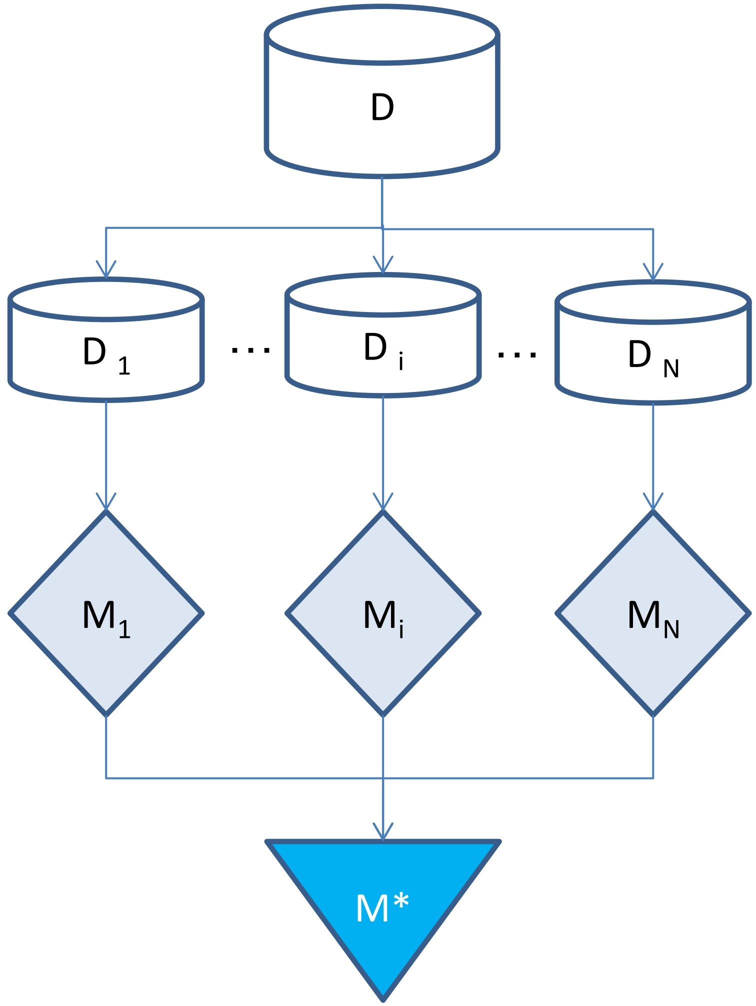
- **Boosting**. Building multiple models (typically of the same type) each of which learns to fix the prediction errors of a prior model in the chain.

- **Stacking**. Building multiple models (typically of differing types) and a supervisor model that learns how to best combine the predictions of the initial models.

Bagging (Bootstrap Aggregating)

Steps	Modeling	Considerations

Start With:
Training Data Set



Step 1:
Create Multiple Sub Sets

Resampling Method?

Step 2:
Build Multiple Models

What type of model(s)?

Step 3:
Combine Models

Method of combing
models or outputs?

Example see code: https://github.com/mshump/models_with_models

Boosting

Steps	Modeling	Considerations

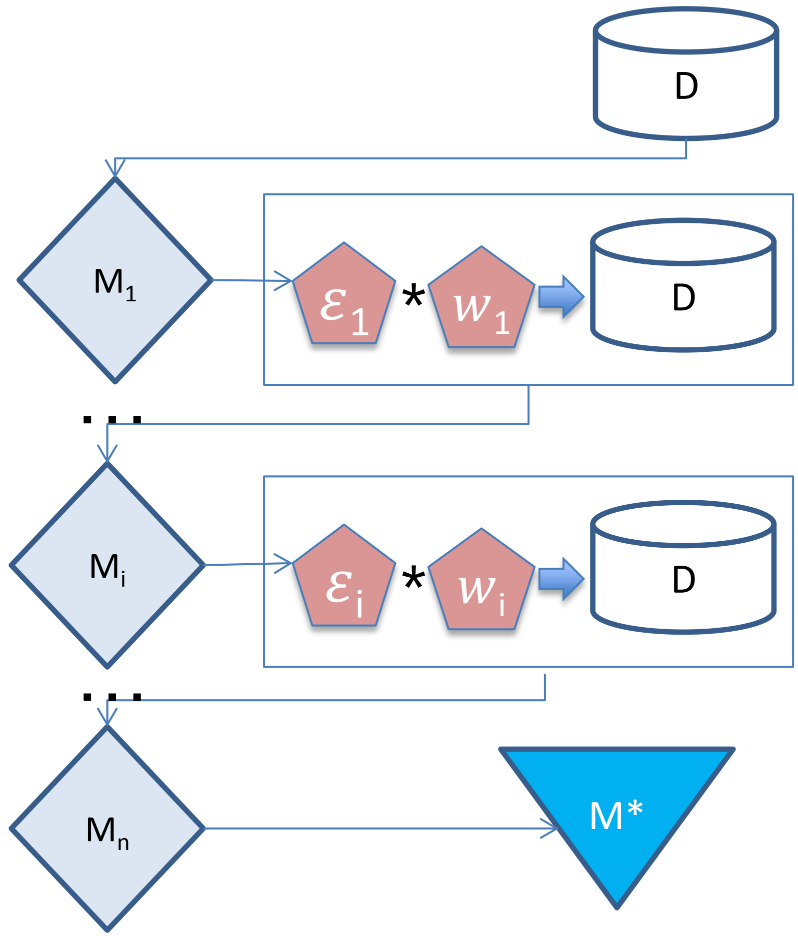
Start With:
Training Data Set

Step 1:
First model is learned on the whole training data set (w/ equal obs weights)

$$Y_1 \sim f(x)_1 + \varepsilon rror_1$$

Step 2:
Following models are learnt on the training set based on the performance of the previous (w/ higher weights on poorly predicted obs)

Step 3:
Continue to add learners until a limit is reached in the number of models or accuracy.



What type of model (same or different)?

What weighting method?

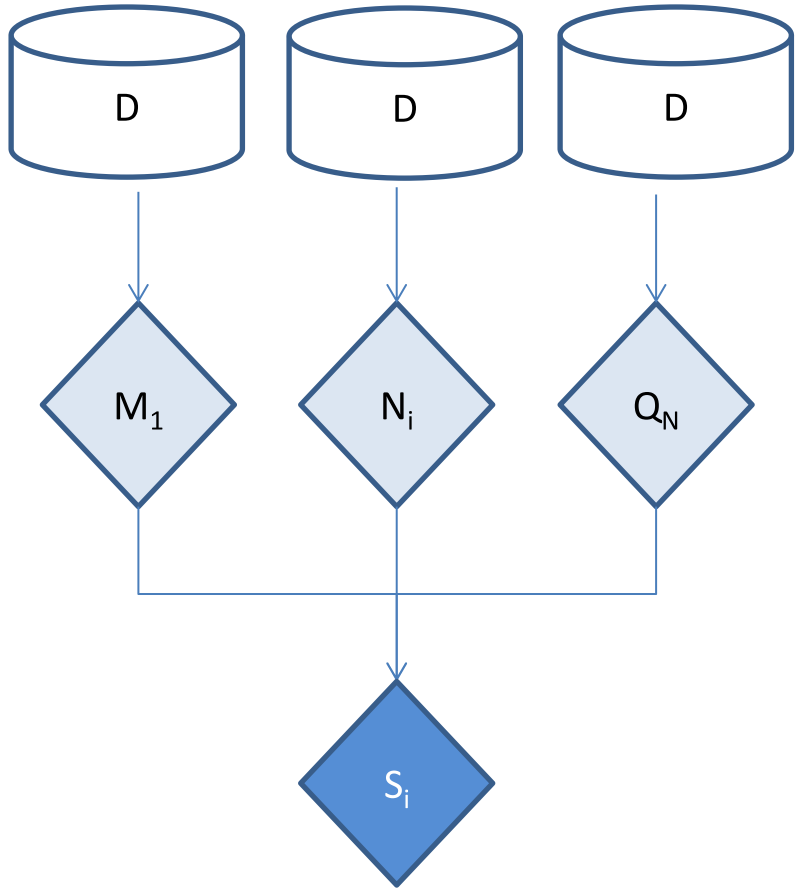
Stopping and combining decisions?

Example see code: https://github.com/mshump/models_with_models

Vertical Stacking

Steps | Modeling | Considerations

Start With:
Training Data Set



Step 1:
Multiple base models (usually different types) to predict on training set

What type of models?

Step 2:
Develop supervisor model / meta learner which best combines predictions from base models

What type of supervisor model?

Example see code: https://github.com/mshump/models_with_models

R Sample Code Output (Example 6)

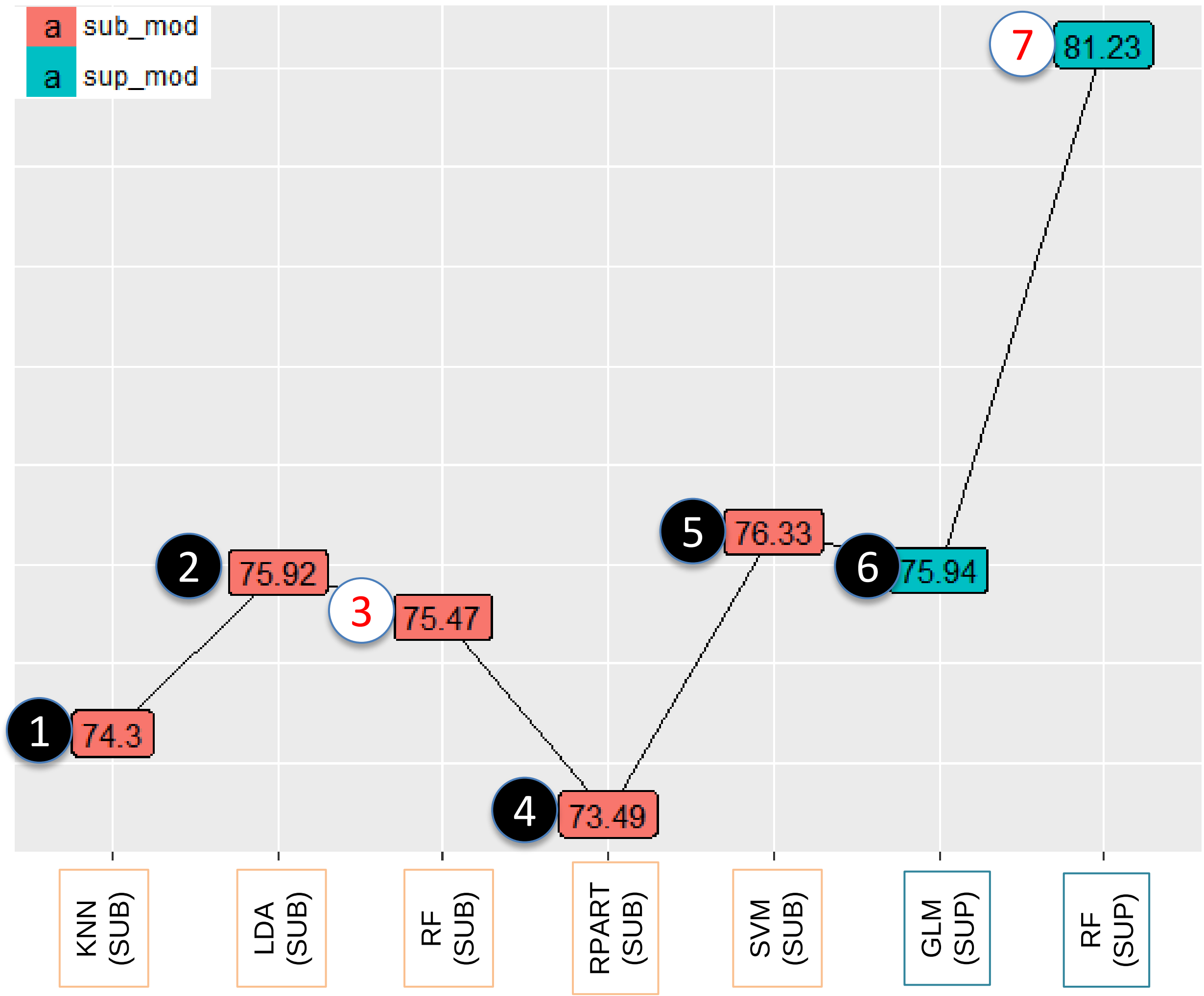
Vertical Stacking - Binomial Classification

5 sub-models learned on sample training data

1. K-Nearest Neighbors
2. Linear Discriminant
3. Random Forest
4. Partition Tree
5. SVM (Radial)

2 supervisor models tested

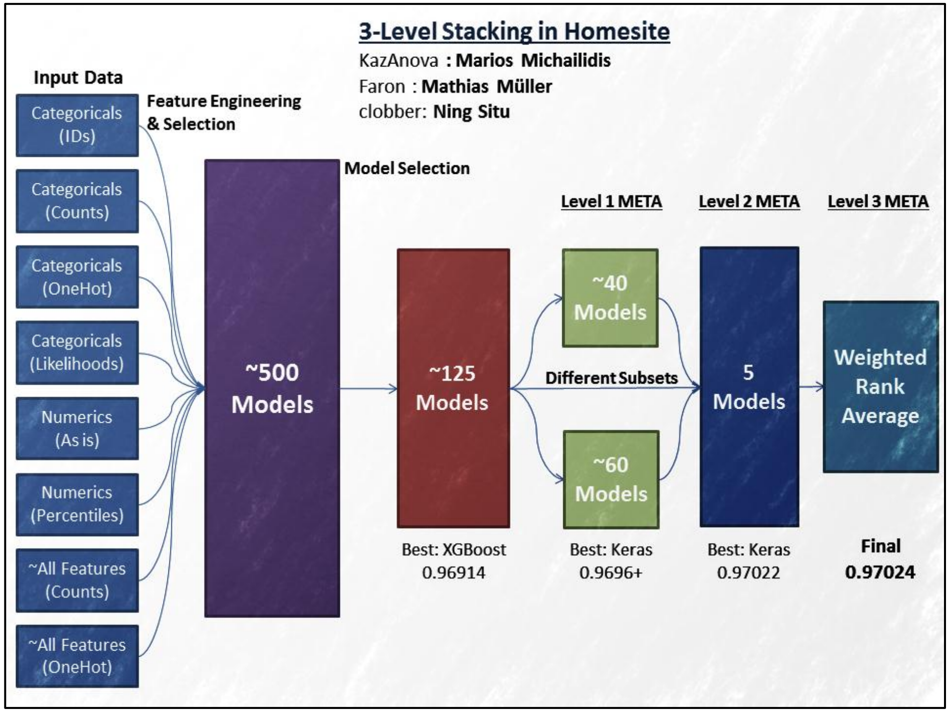
6. GLM - shows no improvement over sub-models
7. Random Forest - show ~5% increase over best sub-model



Pipelines - Some Kaggle Competition Examples

Homeside Quote Conversion:

Which customers will purchase an insurance plan?



Truly Native:

Predict which web pages served by StumbleUpon are sponsored