Stacking

BY MG ANALYTICS

When to choose what??

- **Decrease**
  - **Variance** (bagging)
  - **Bias** (boosting)
  - **Improve predictions** (stacking).

Stacking

- **Stacked** Generalization or "**Stacking**" for short is an ensemble machine learning algorithm.

- combining the predictions from multiple machine learning **models** on the same dataset.

- to explore a space of different models for the same problem.

- Learning with different types of models which are capable to learn some part of the problem, but not the whole space of the problem.

- Models to be stacked should be non linear in nature.
- The models are combine using a linear method.
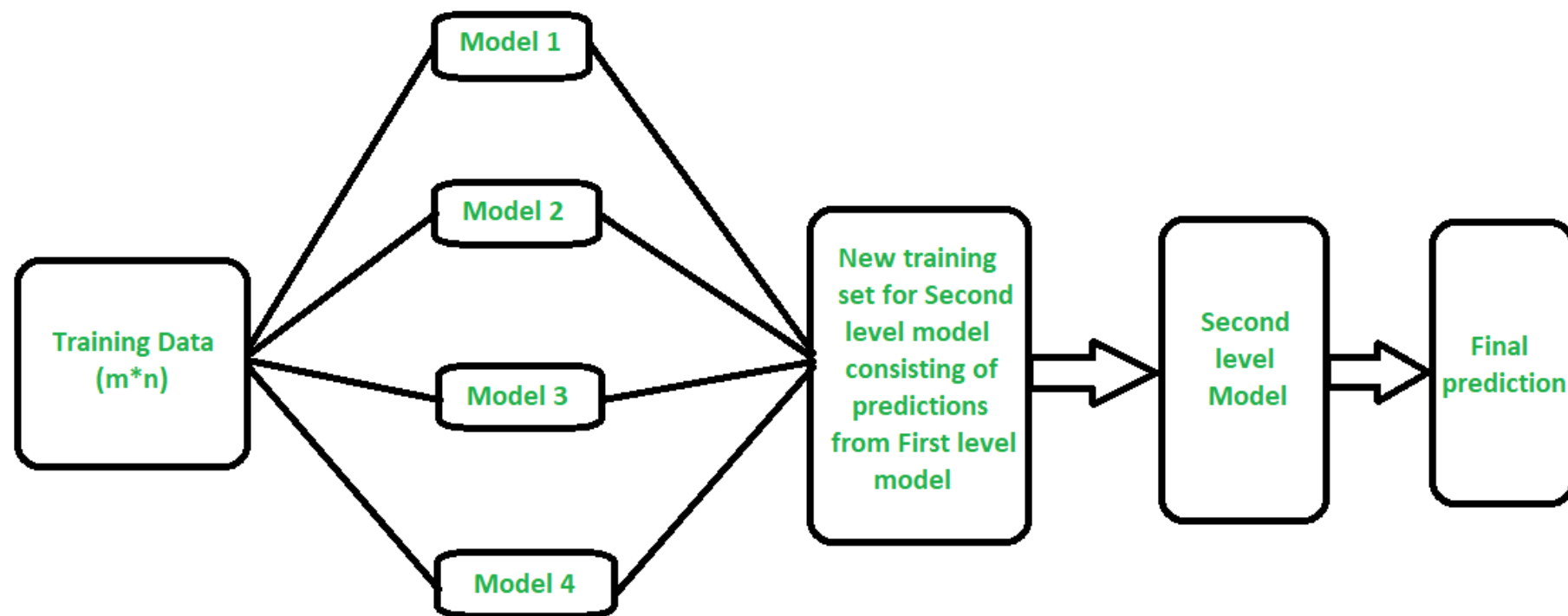
Process

- Initial training data (**X**) has *m* observations, and n features (so it is m x n).

- There are M different models that are trained on X (by some method of training, like cross-validation) before hand.

- Each model provides predictions for the outcome (y) which are then cast into a second level training data (Xl2) which is now m x M. Namely, the M predictions become features for this second level data.

- A second level model (or models) can then be trained on this data to produce the final outcomes which will be used for predictions.

Cross Sampling

- Stacking uses a similar idea to k-folds cross validation to create **out-of-sample** predictions.

- if we were to use predictions from the M models that are **fit to all the training data**, then the second level model will be biased towards the best of M models.

- We alternately train the Models on k-1 folds and predict the out of box kth fold using this trained model.

- This allows the predictions to be out of sample and useful for training of next layer.