



ITD105 – BIG DATA ANALYTICS LECTURE #7

PAUL B. BOKINGKITO JR.



Improve Performance with Ensembles



What you will learn



Improve Performance with
Ensembles

VOTING

Multiple Heterogenous Models into One Prediction

BAGGING

Multiple Homogenous Models in to One Prediction
(Parallel)

BOOSTING

Multiple Homogenous Models into One Prediction
(Sequential)

Outline

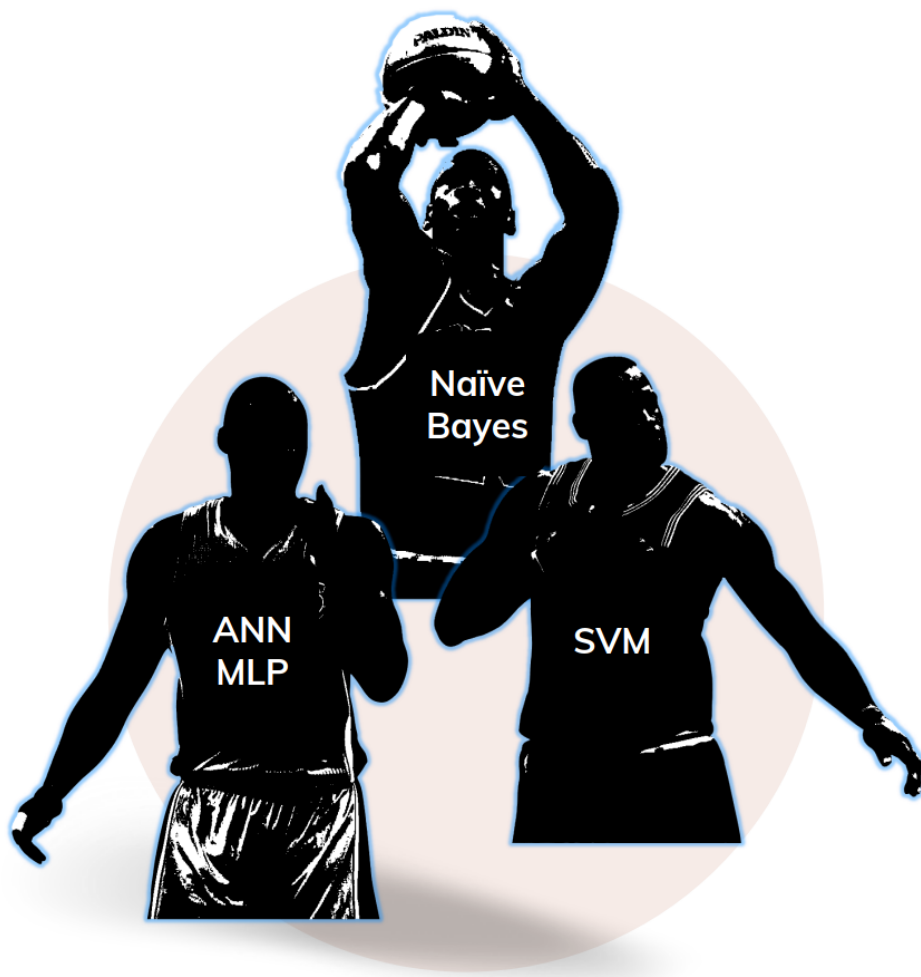
Points for discussion

Ensemble

Voting

Bagging

Boosting



Select Best Model

You have trained models with parametric and non-parametric ML algorithms.

So far, what **model** has the **highest accuracy**?

Instead of relying on one, what if we **combine** results?



+



+



= **ENSEMBLES**

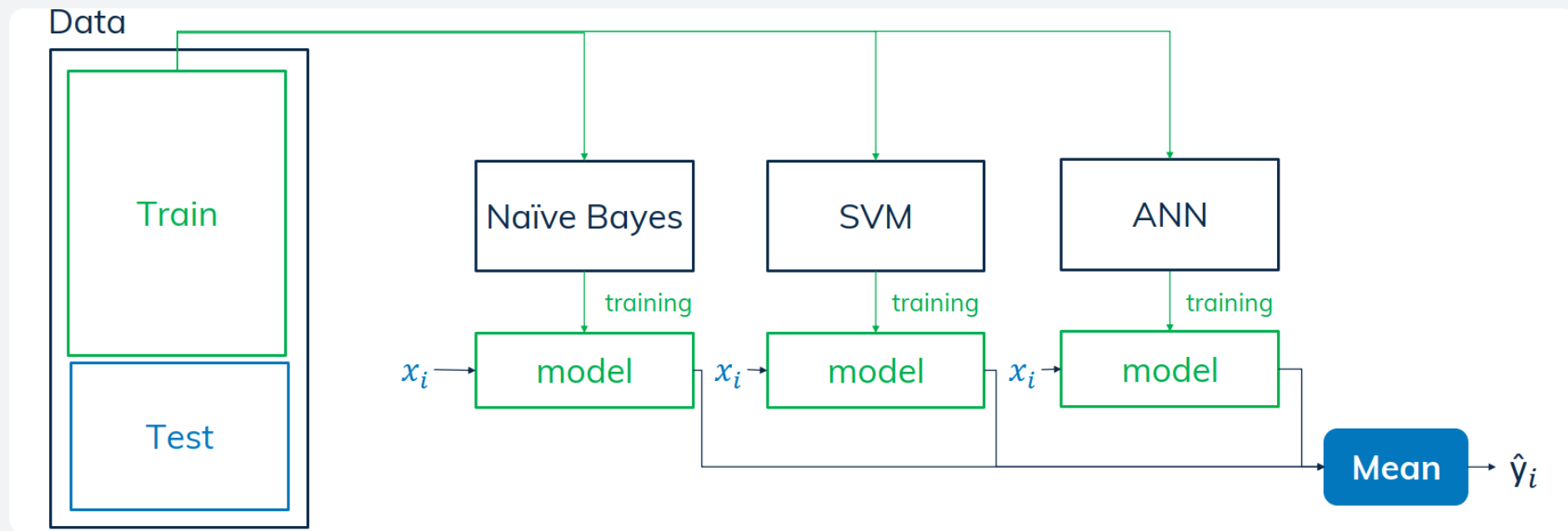
Three most popular methods

Combine Models into Ensemble Predictions

- **Voting.** Building multiple models (typically of different types) and simple statistics (like calculating the mean) are used to combine predictions.
- **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 sequence of models.

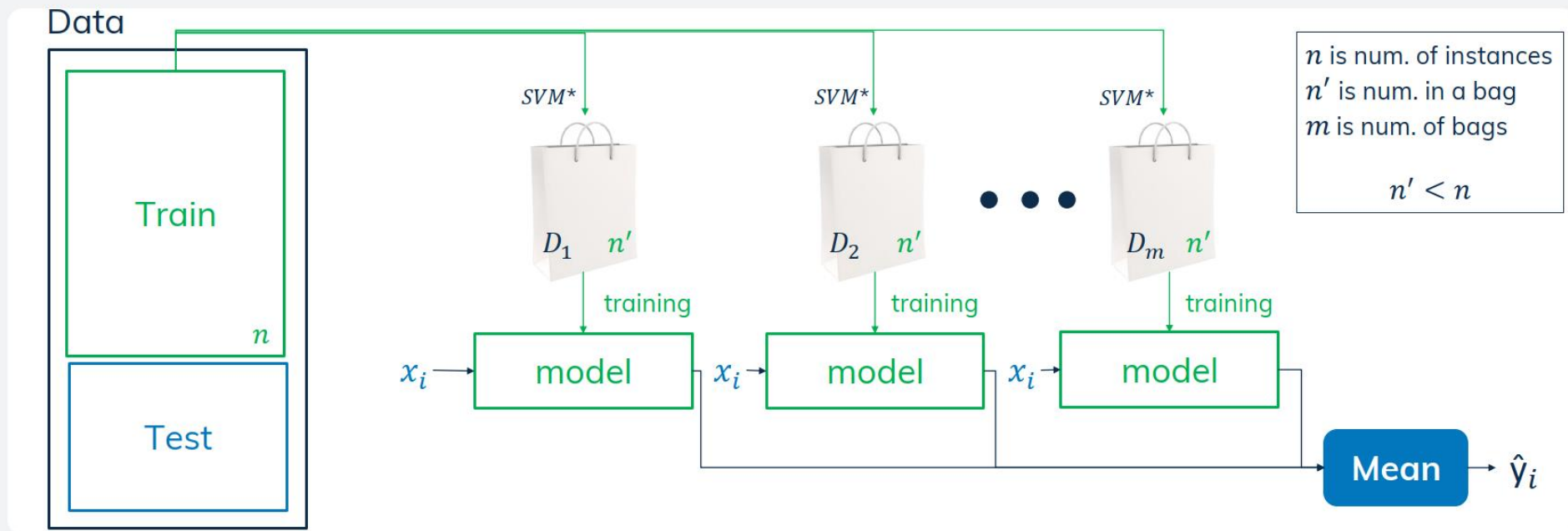
VOTING

Combine Heterogenous Models into One Ensemble Prediction



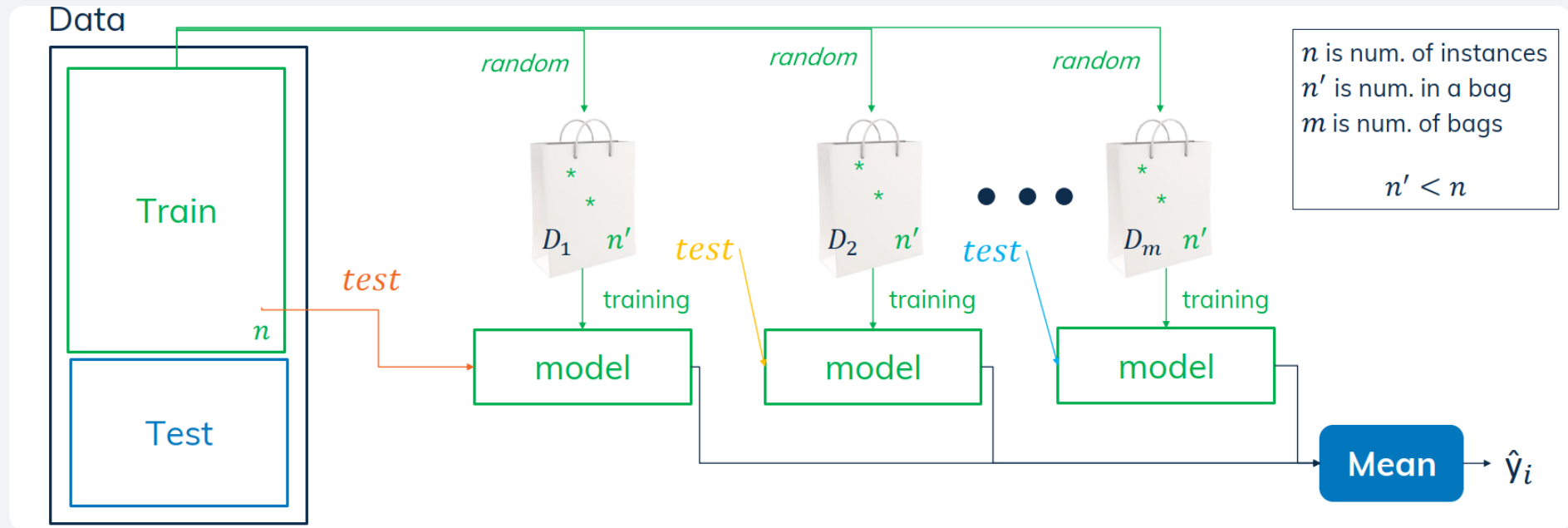
BAGGING (Bootstrap Aggregating)

Combine Homogenous Models into One Ensemble Prediction (Parallel)

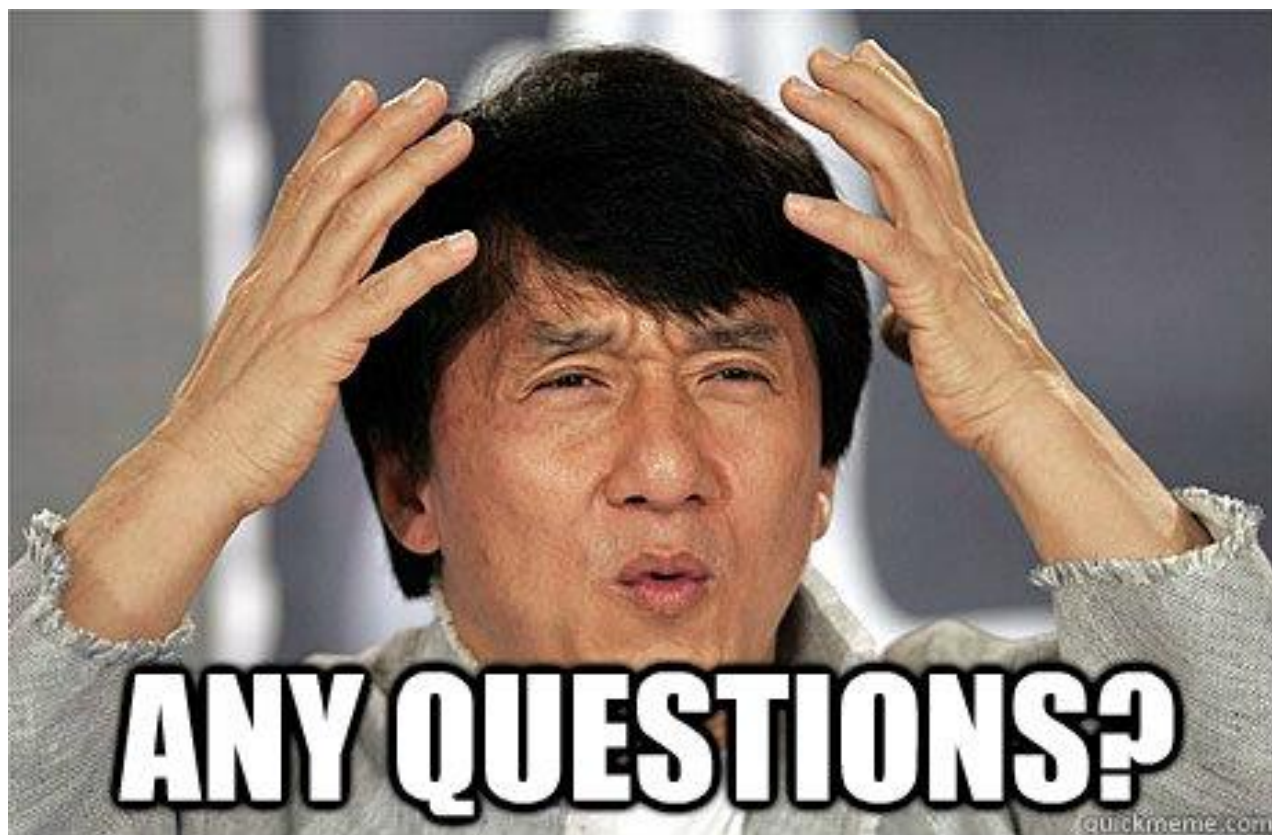


Boosting

Combine Homogenous Models into One Ensemble Prediction (Sequential)



DEMO



END