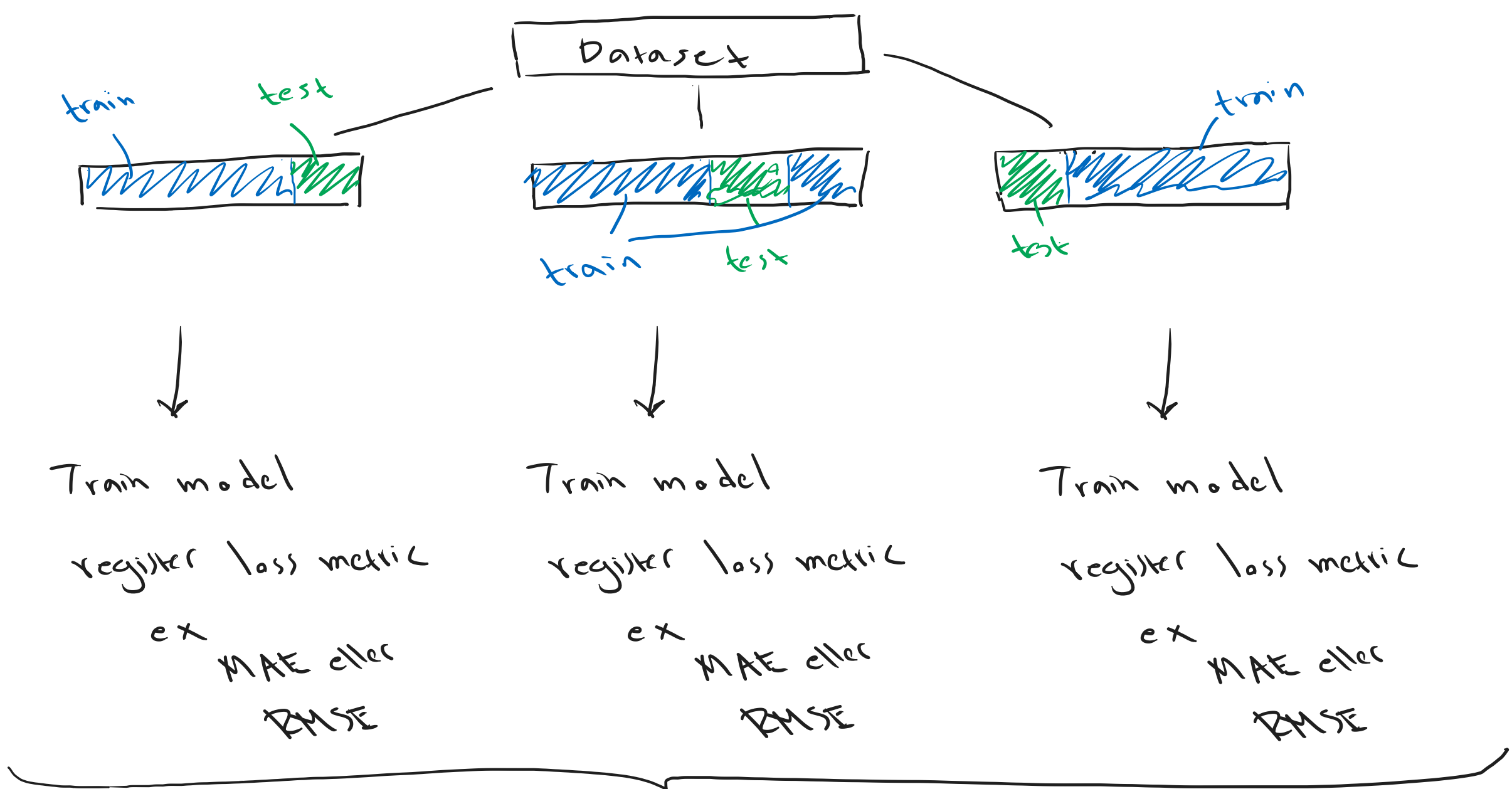


Cross validation

Cross validation används för att undvika tur/otur som kan uppstå när man splittar sitt dataset till train/test-splits. Ibland kan nämligen denna split ske på ett sånt sätt som ex. gör det idealt för en overfittad modell att prestera bra på test-setet.

Låt oss demonstrera Cross Validation nedan.

3-fold cross validation



Nu har vi tränat en viss modell, säg en linjär modell, på exakt samma features men på i detta fall tre olika train/test splits. Vi har också registrerat modellernas performance metrics (dvs ex MAE).

Vi säger nu att det bästa estimatet av sann prestanda ges av medelvärdet av alla dessa. Dvs,

I detta fall

$$MAE_{total} = \frac{MAE_1 + MAE_2 + MAE_3}{3}$$

Cross validation används alltid för kombinationen modell + features