

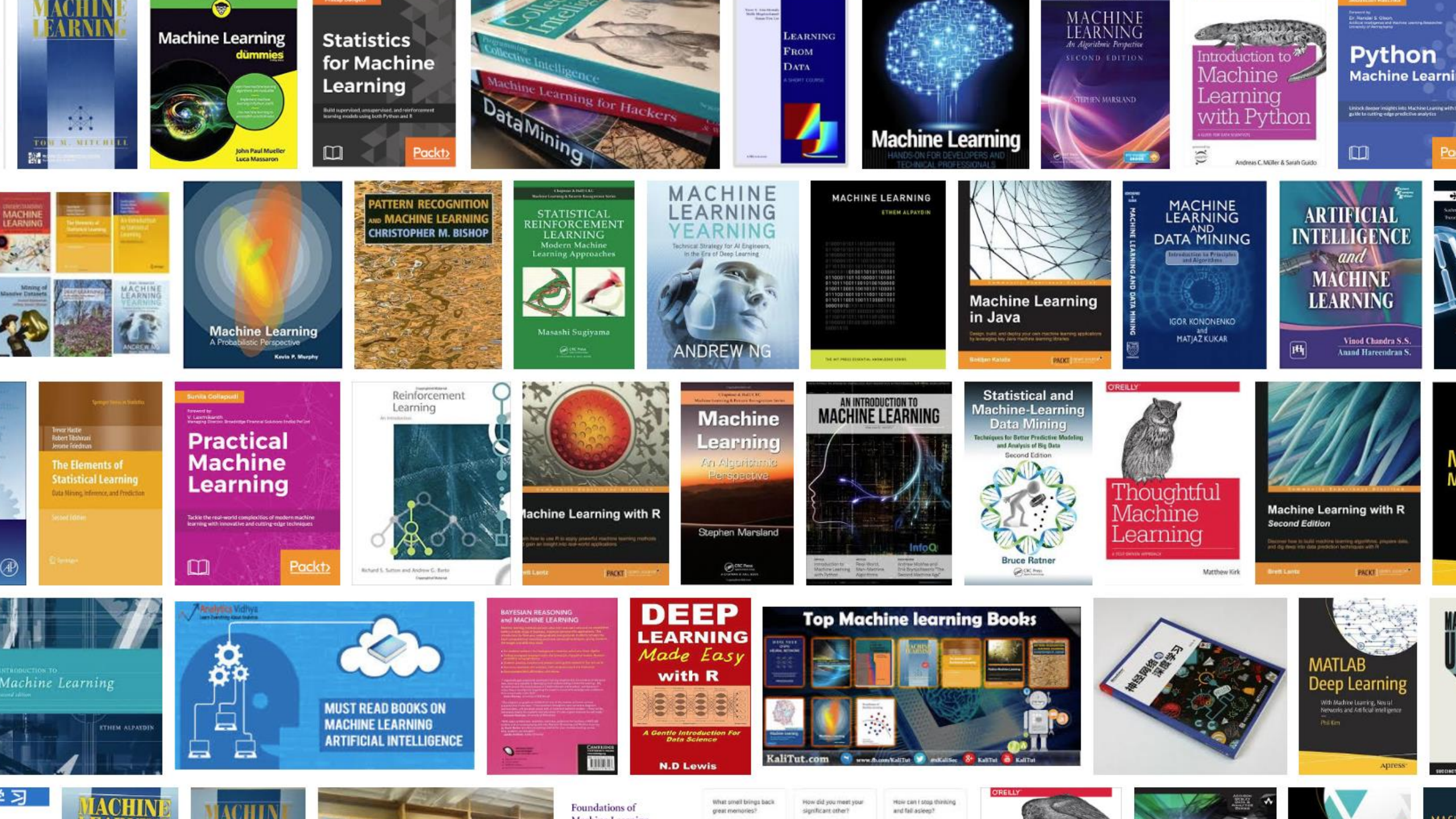
# Managing multiple models in production

Terry McCann | @SQLShark

Principal Data Scientist





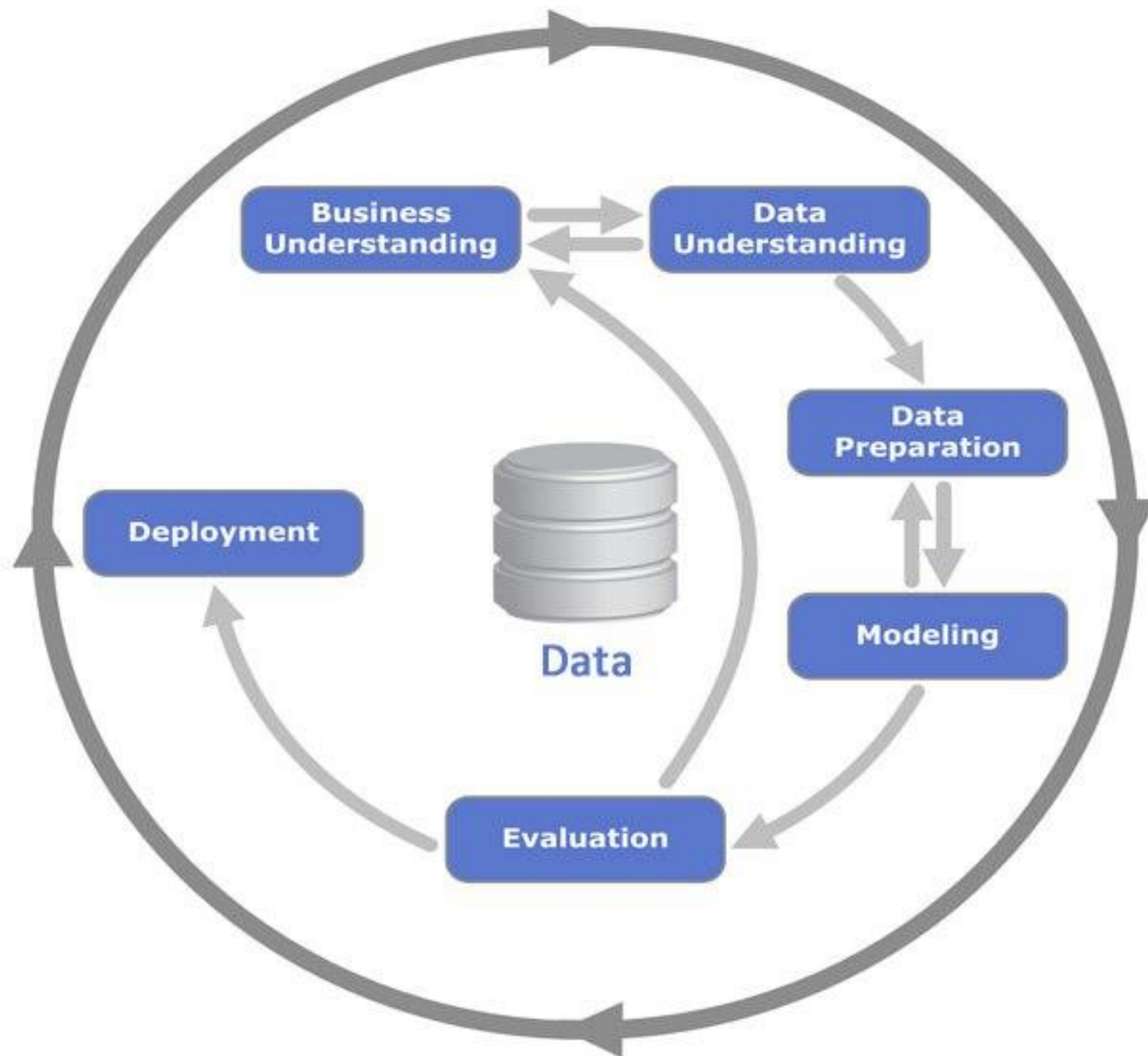


# The Machine Learning Process

1. Gathering data
2. Preparing that data
3. Choosing a model
4. Training
5. Evaluation
6. Hyperparameter tuning
7. Prediction
8. Deployment



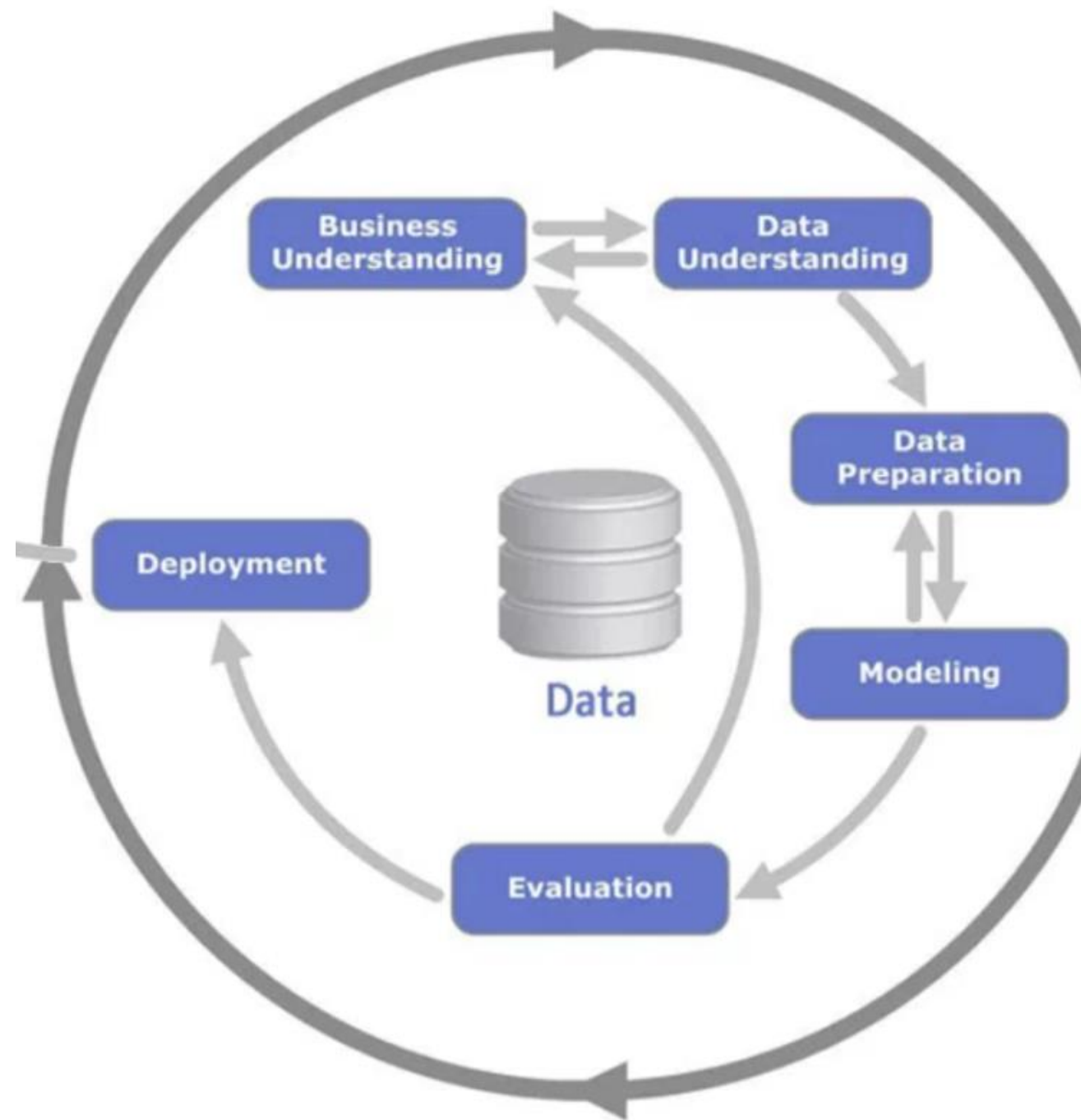
# CRISP-DM Process Diagram





“Productionisation of models is the ***TOUGHEST*** problem in data science”

What is wrong with the development process?

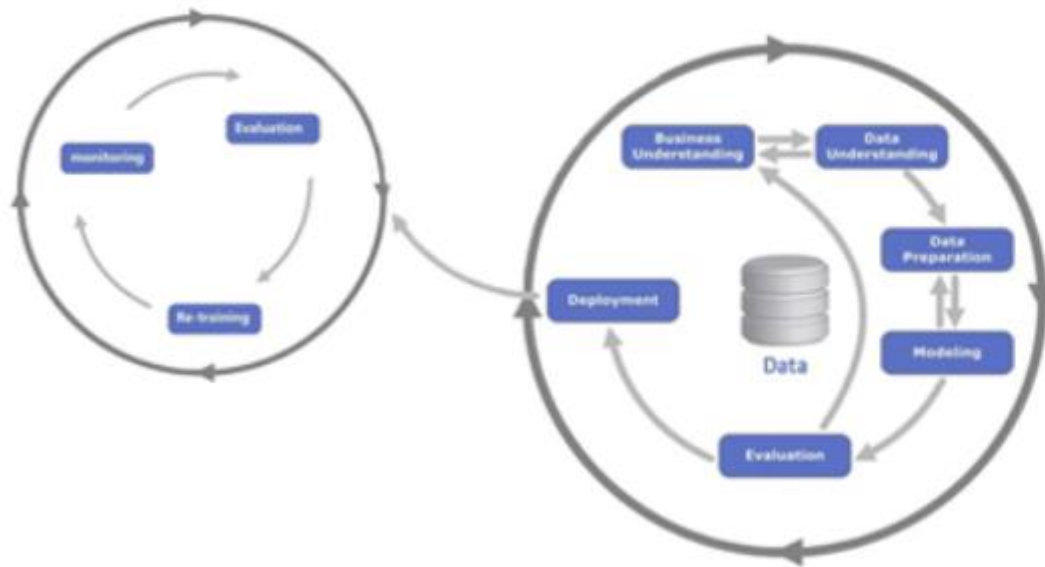




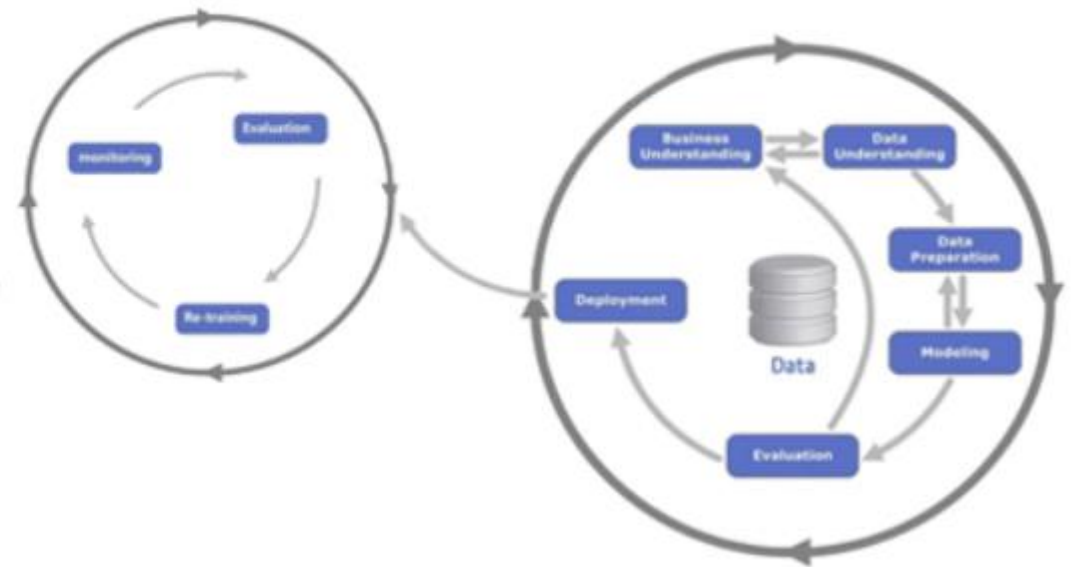


# AB Testing scenario

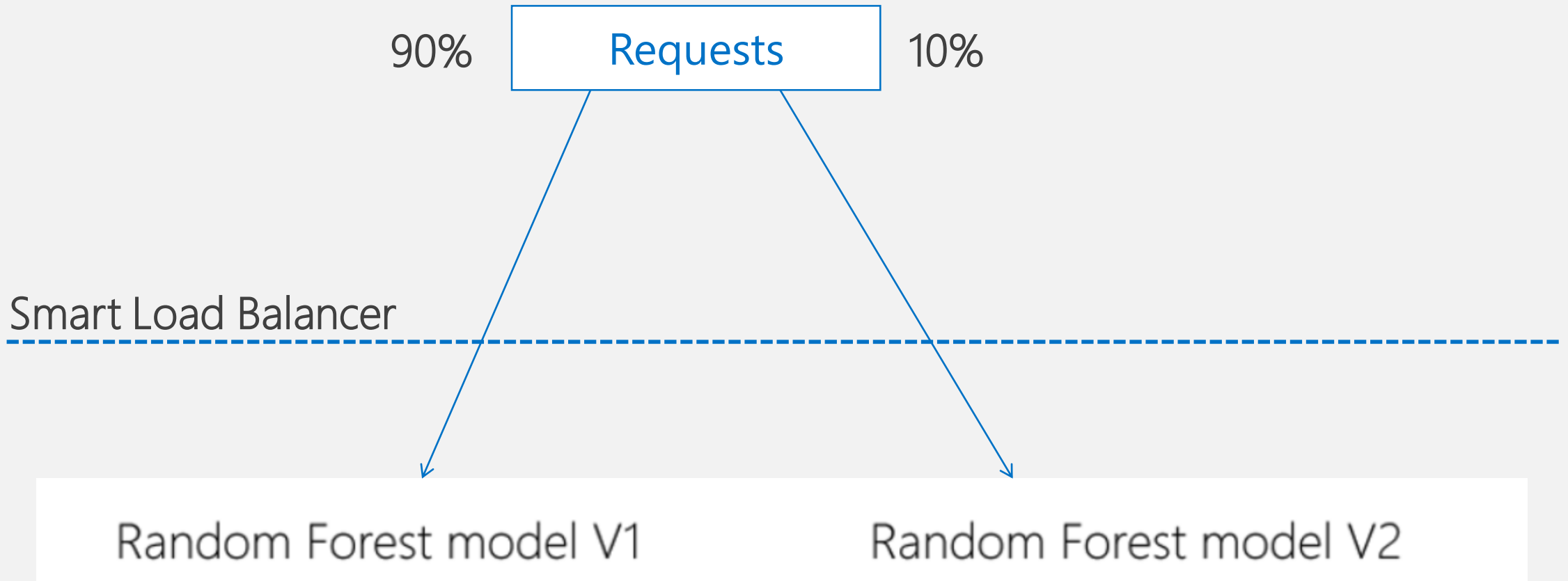
Random Forest model V1



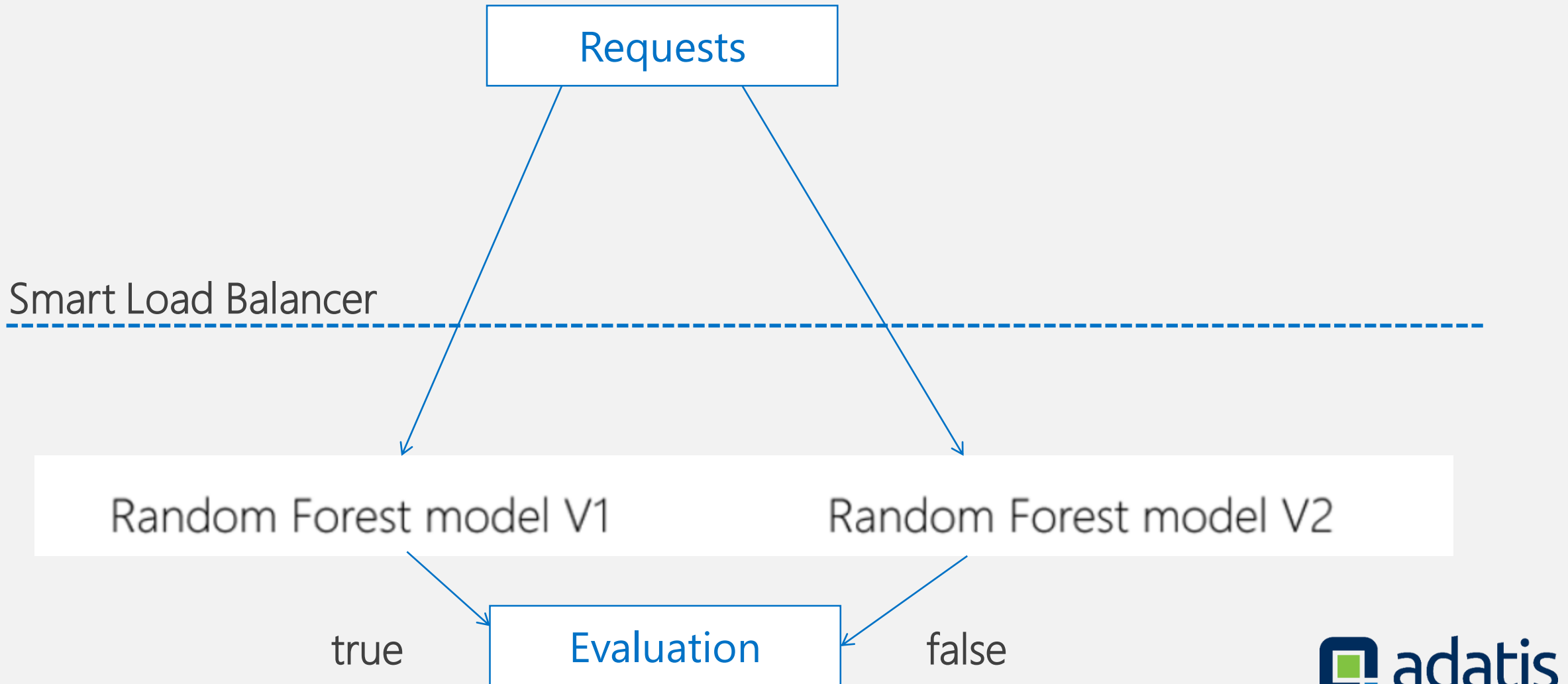
Random Forest model V2



# AB Testing scenario

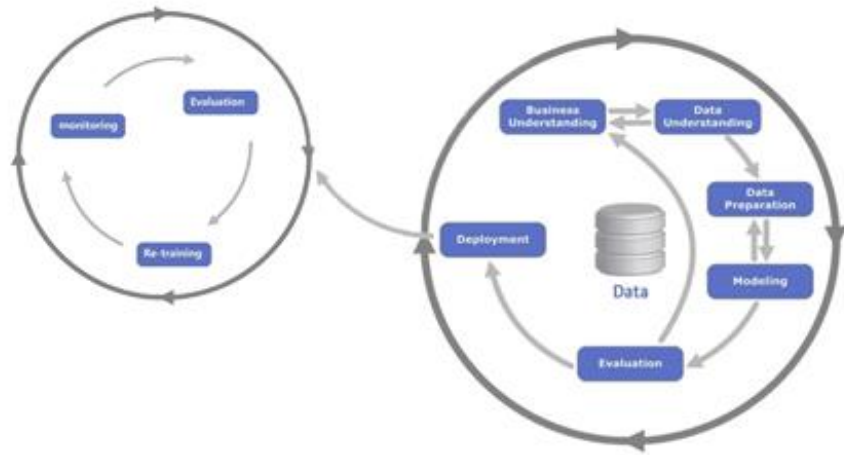


# AB Testing scenario

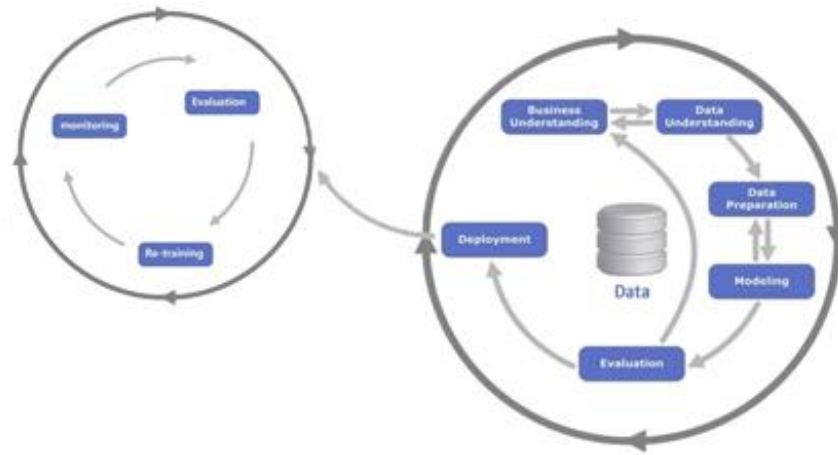




Random Forest model V1



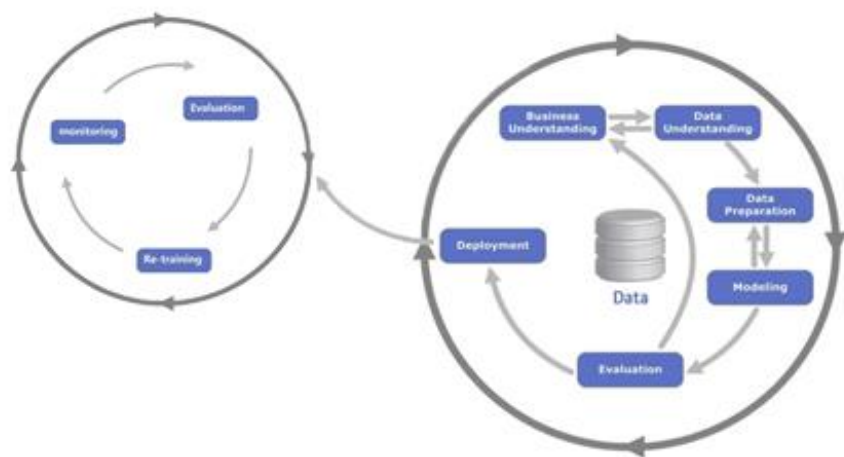
Random Forest model V2



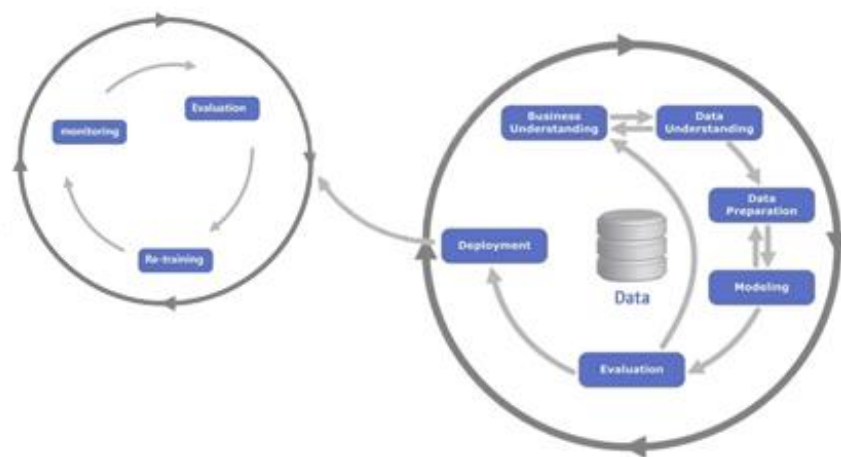
A Smart Load Balancer, just wont work.

But this isn't the only  
problem...

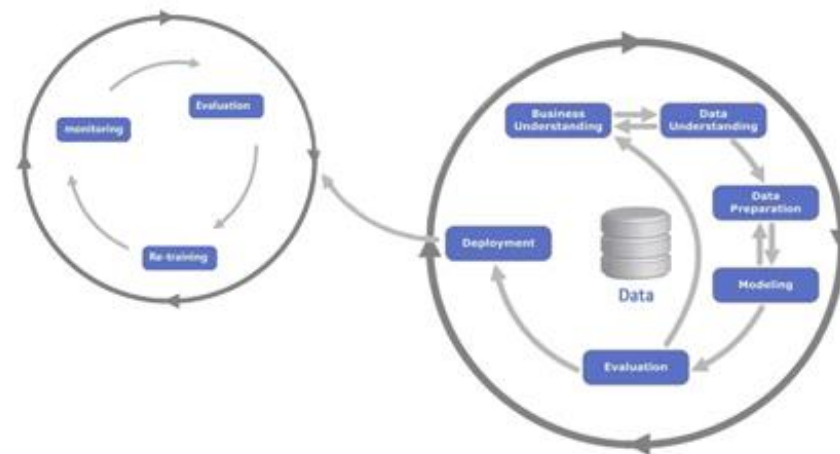
Random Forest model V1



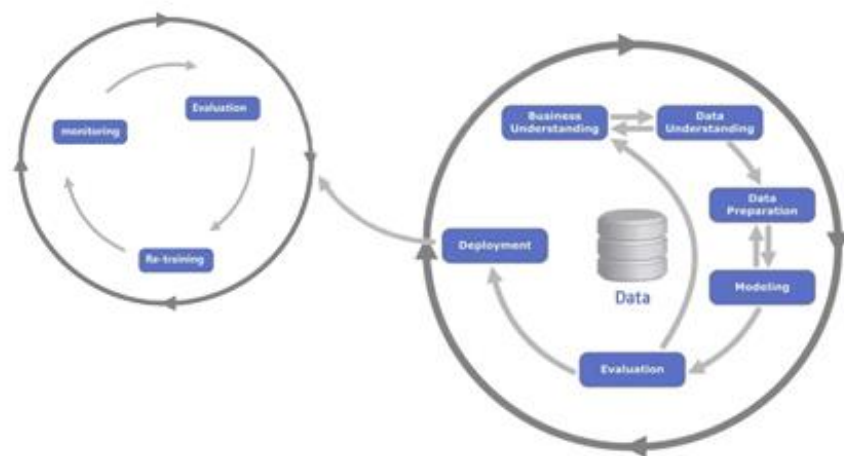
Random Forest model V2



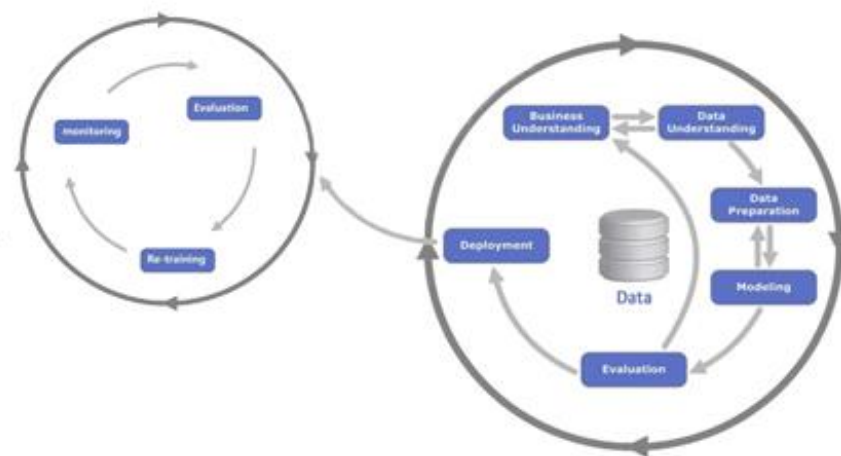
K-Nearest Neighbours



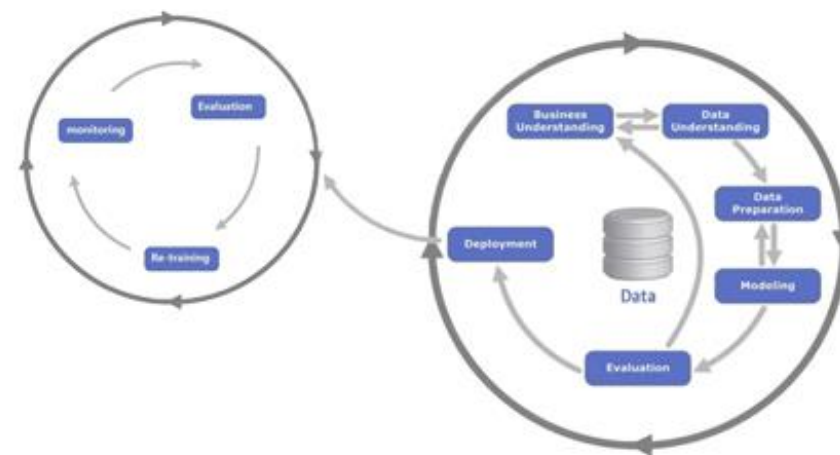
Ada Boost



Naive Bayes

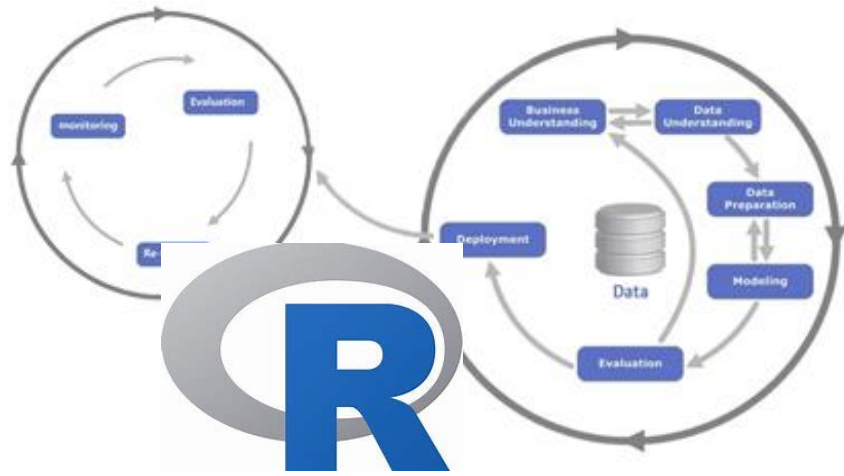


Decision Tree

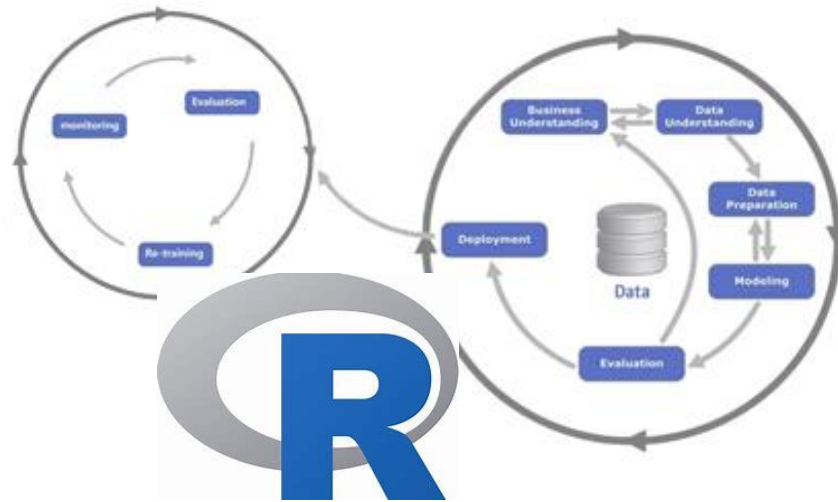




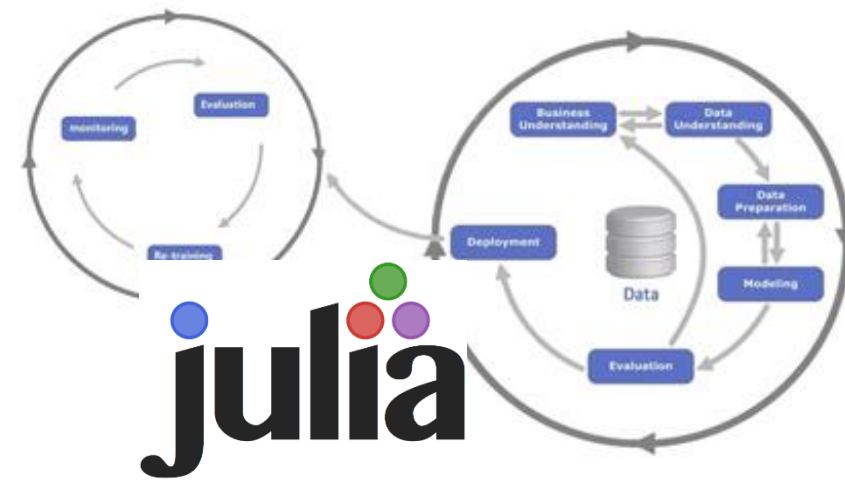
Random Forest model V1



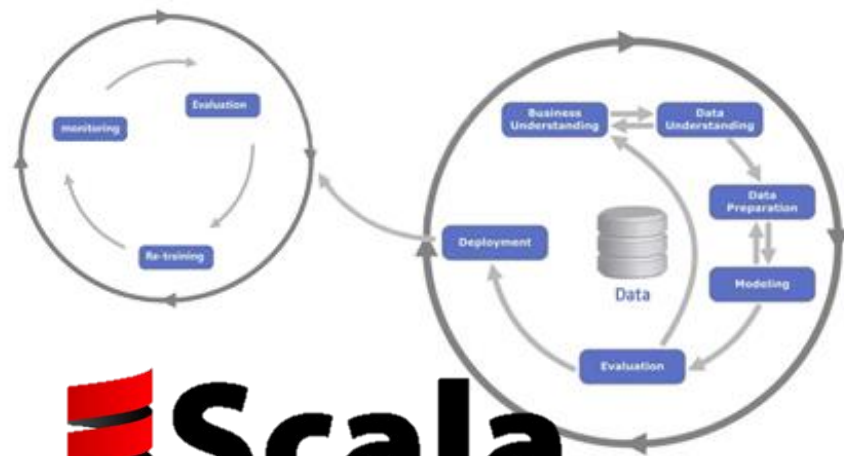
Random Forest model V2



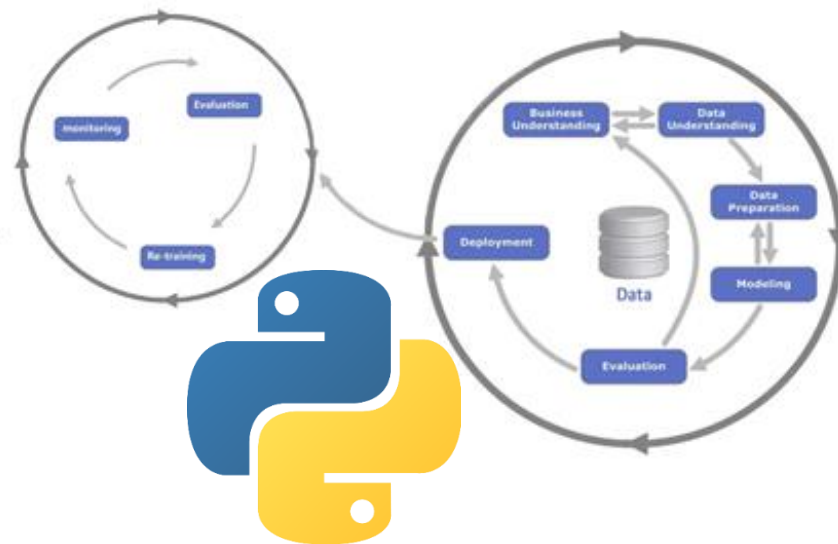
K-Nearest Neighbours



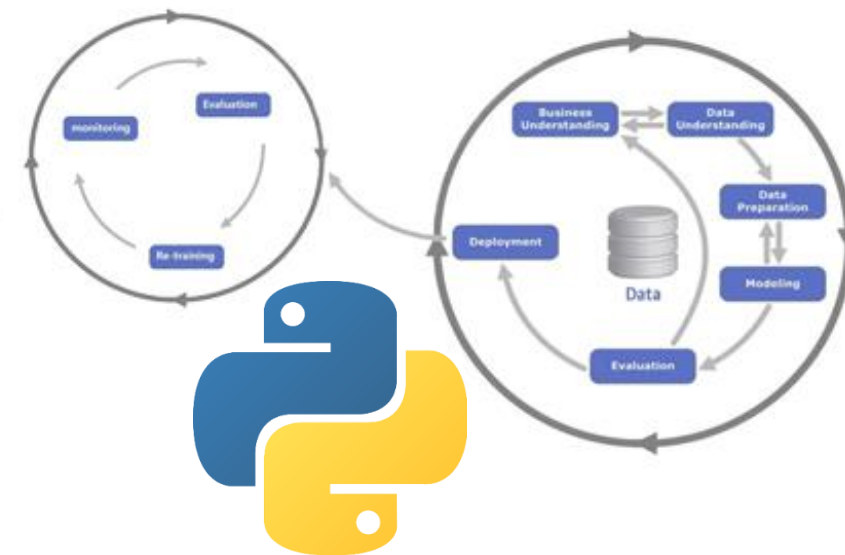
Ada Boost



Naive Bayes



Decision Tree



 **Scala**

Smallest company 5

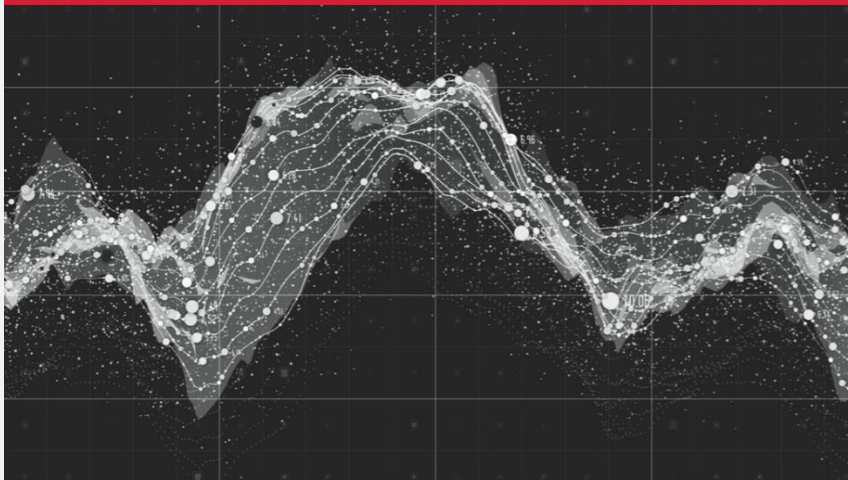
Largest company 13

Where do you start?

O'REILLY®

# Machine Learning Logistics

Model Management in the Real World



Ted Dunning & Ellen Friedman



Ted Dunning

<https://mapr.com/blog/author/ted-dunning/>

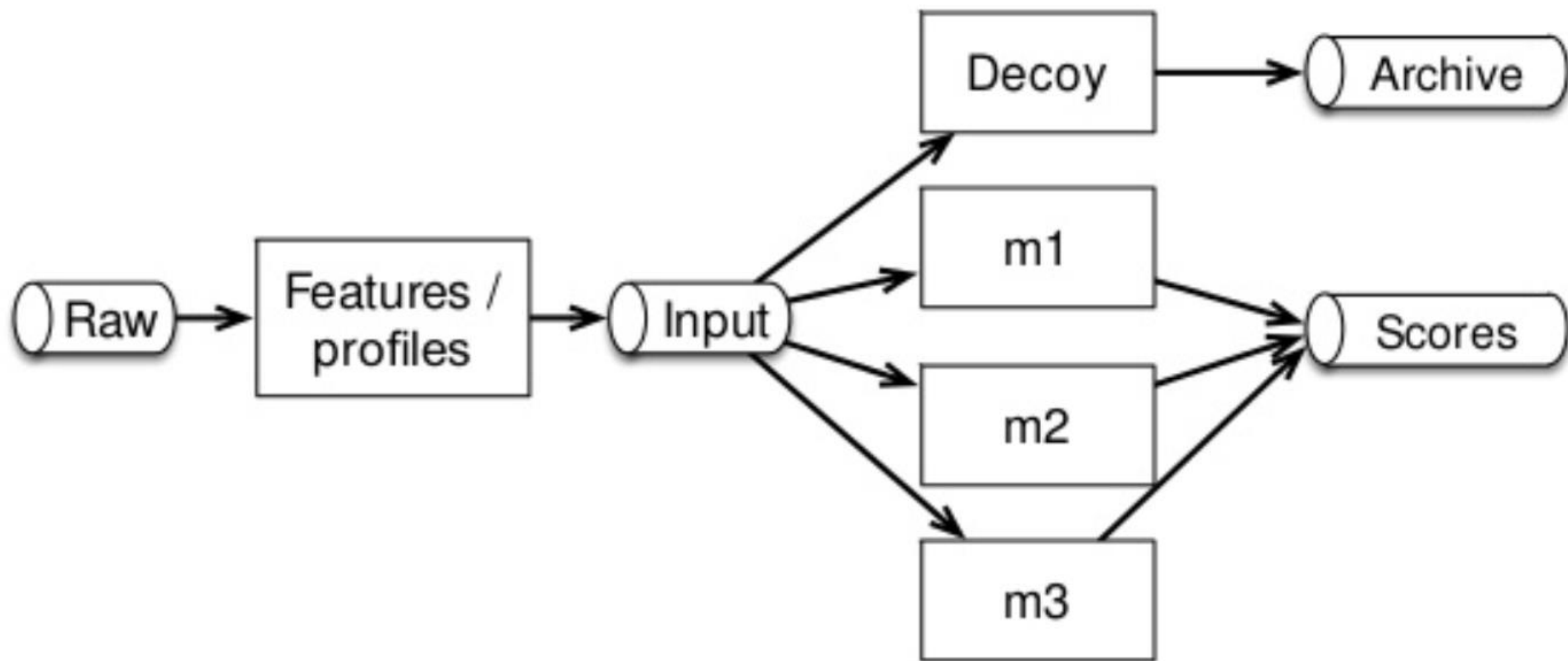
Ellen Friedman

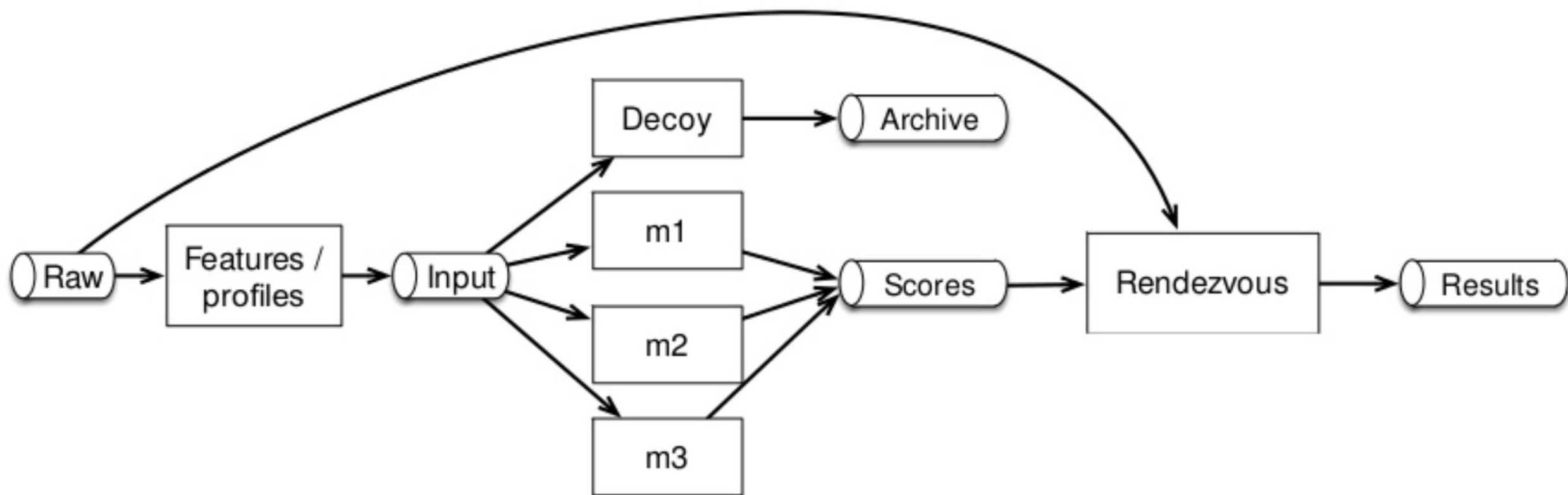
<https://mapr.com/blog/author/ellen-friedman/>

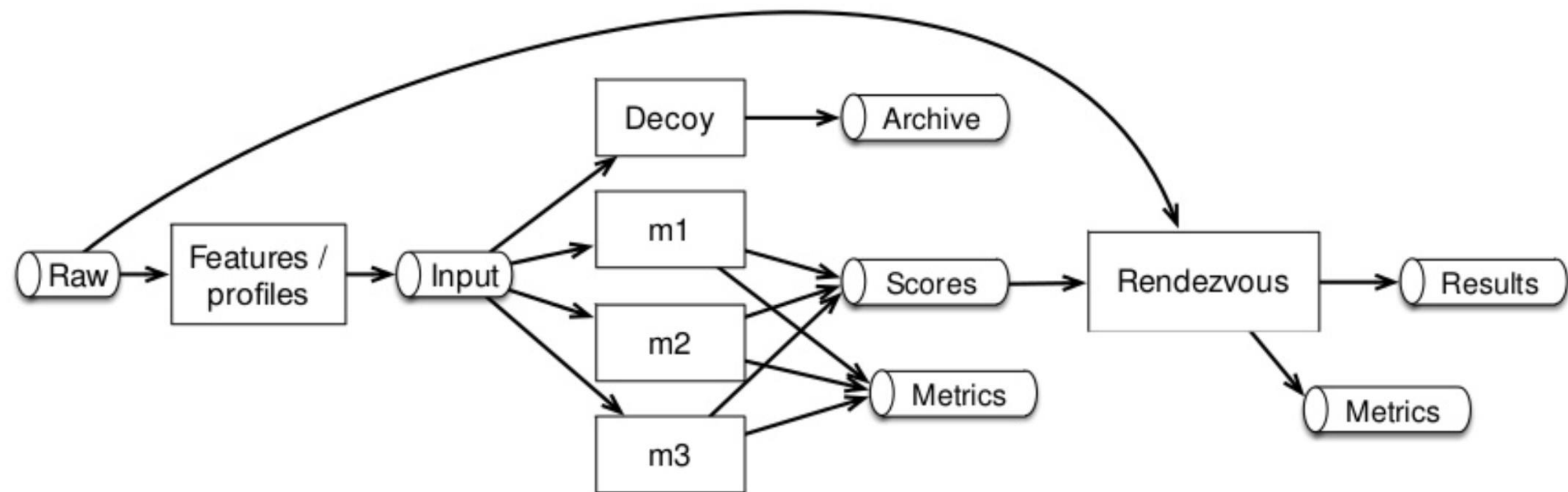




# *Rendezvous Architecture*

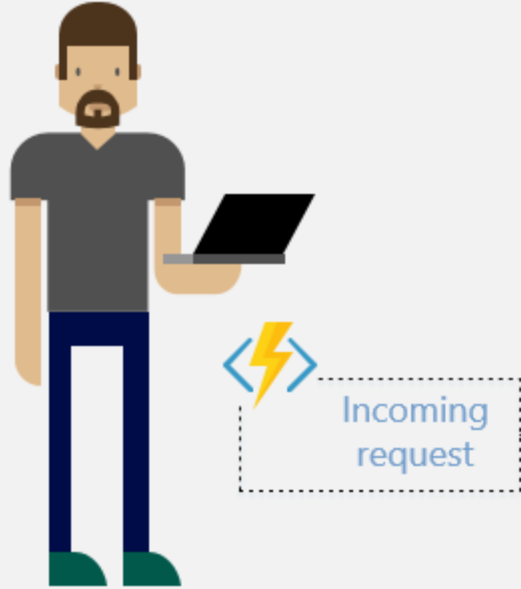




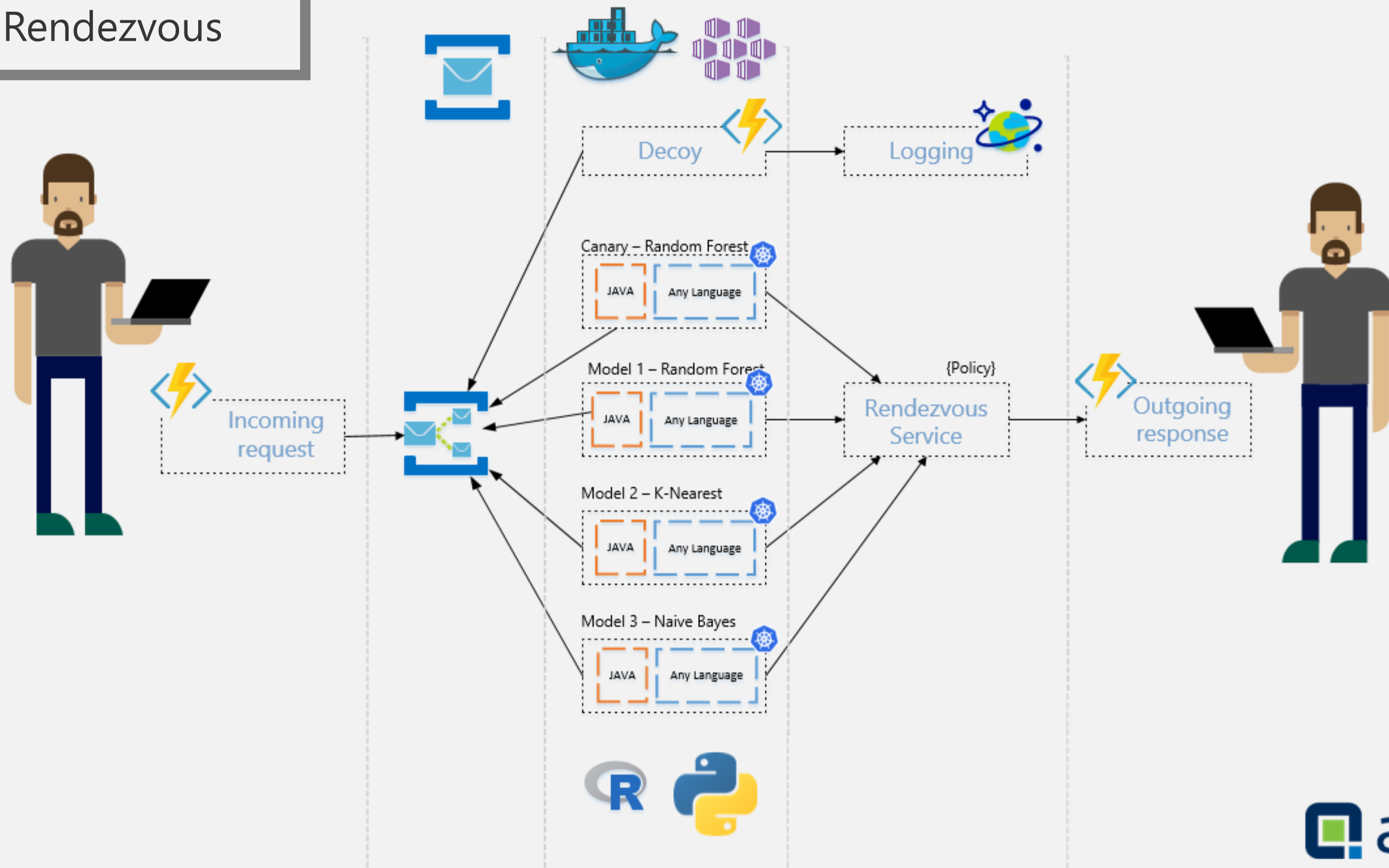


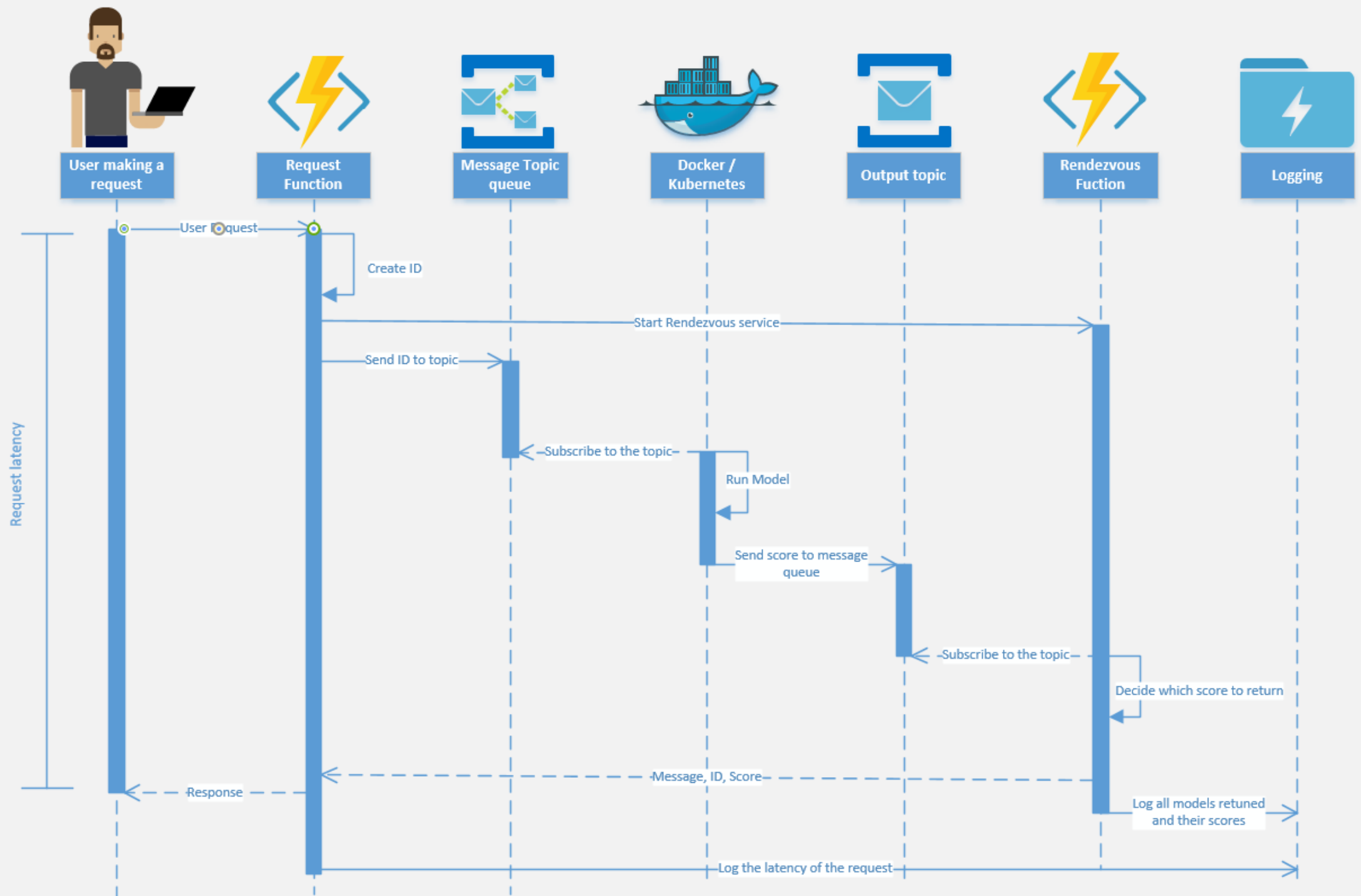


# Rendezvous



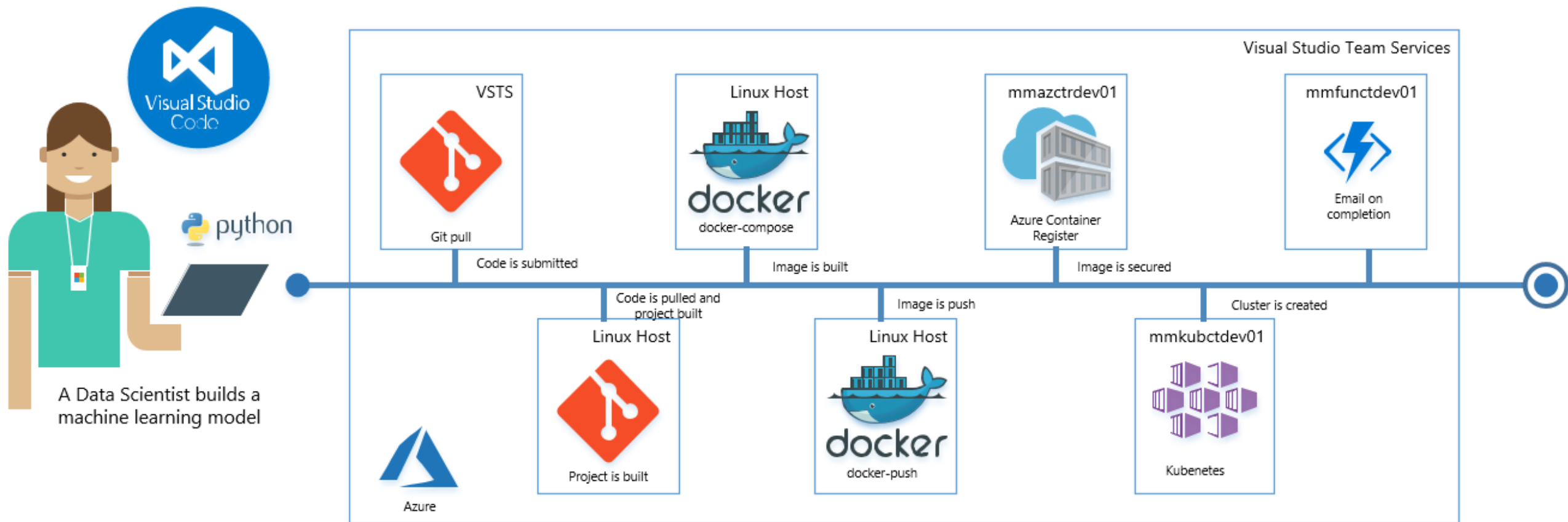
# Rendezvous





Do our Data Scientist's really need to  
learn all of that?

# DataOps



# Managing multiple models in production

Terry McCann | @SQLShark

Principal Data Scientist

Terry.mccann@adatis.co.uk

