



On TAP:

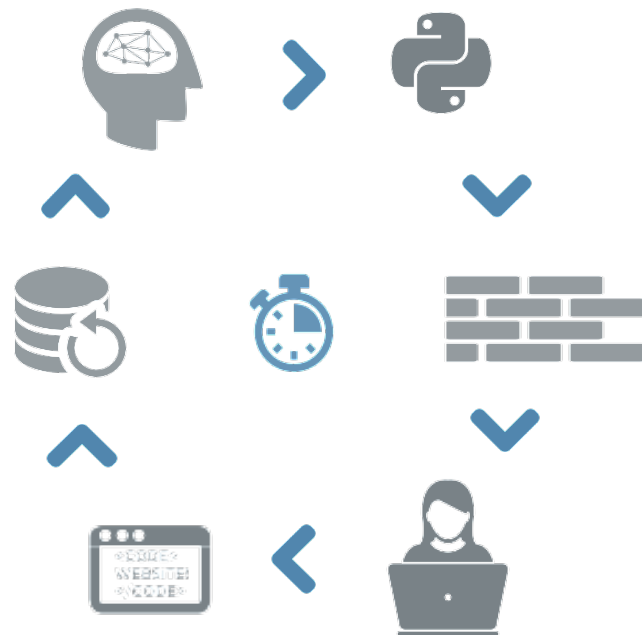
Module 4: Model Building

Kyle H. Ambert, PhD

Intel Big Data Solutions, Datacenter Group



Model Building



Model Building

- Annotate a data frame with gold-standard labels



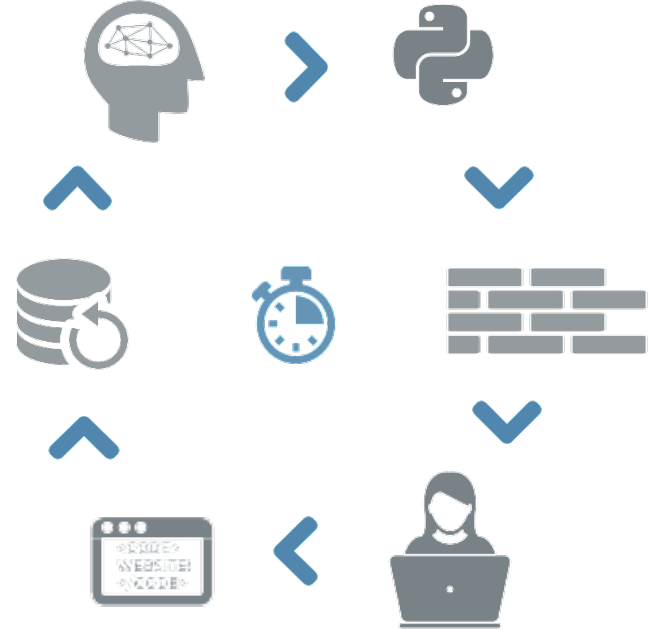
Model Building

- Annotate a data frame with gold-standard labels
- Split data into train and test partitions



Model Building

- Annotate a data frame with gold-standard labels
- Split data into train and test partitions
- Train and Evaluate a random forest model



Problem:

- [1] Use a feature besides admit/discharge date to create a gold-standard label column (i.e., generate “POSITIVE” and “NEGATIVE” labels according to some reproducible rule).
- [2] Build a Random Forest classifier using any combination of features you choose to classify on the gold standard label chosen in [1]. Evaluate.
- [3] What steps would be involved in creating a function to perform cross-validation? Can you think of a way to implement this in TAP?