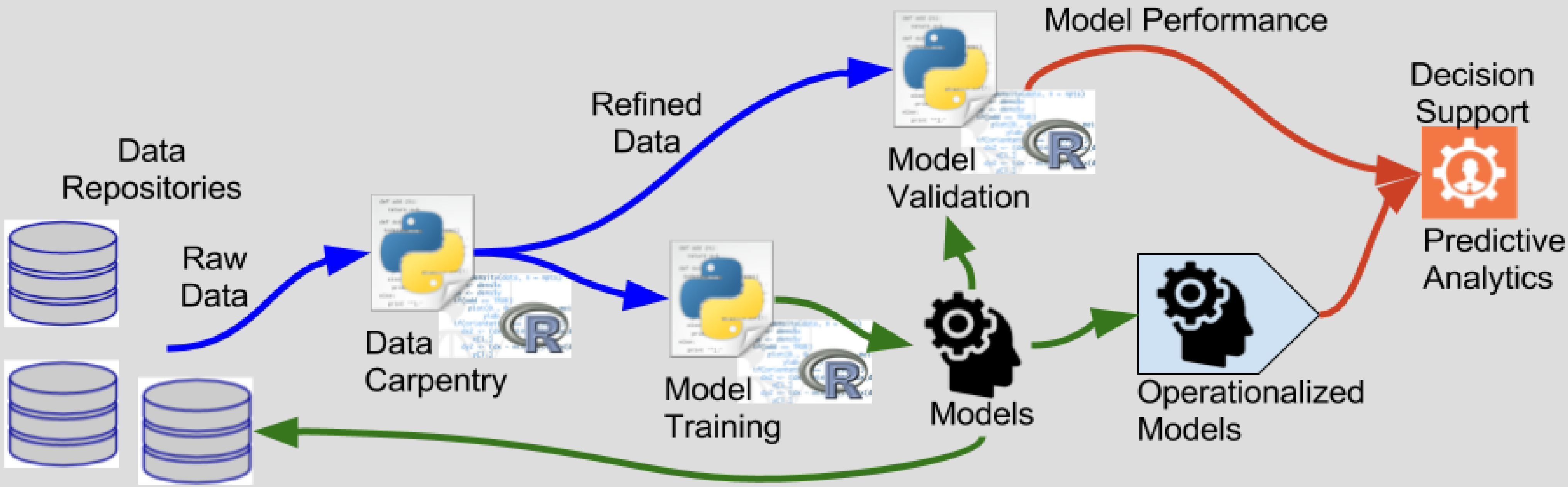
Machine Learning Model Metrics

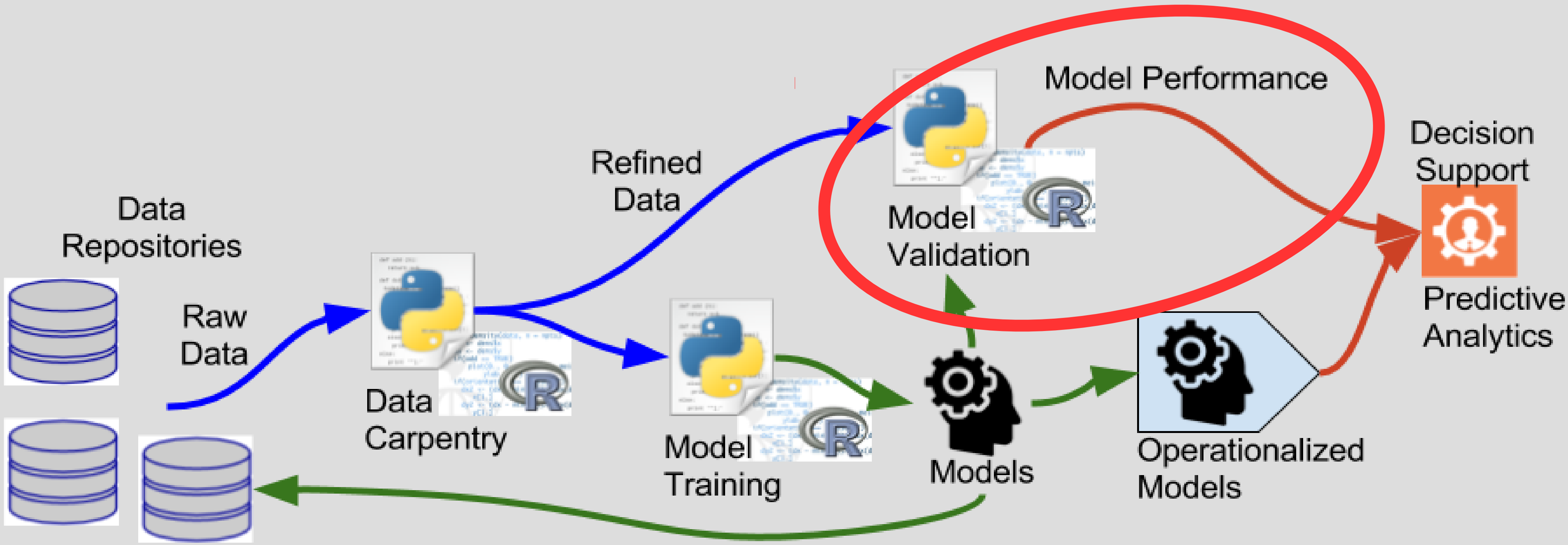
Topics

- Regression Metrics
- Classifier Metrics
- Clustering Metrics

Machine Learning Workflows



Machine Learning Workflows



Regression Metric

- R-Squared (R^2)
  $$R^2 \equiv 1 - \frac{SS_{\text{res}}}{SS_{\text{tot}}}.$$
  - Coefficient of determination
- Adjusted R-Squared
  $$\bar{R}^2 = 1 - (1 - R^2)\frac{n-1}{n-p-1} = R^2 - (1 - R^2)\frac{p}{n-p-1}$$
  - Always less than R^2
  - R^2 can artificially increase with more explanatory variables (independent, predictors, input)
  - Adj. R^2 only increases when the R^2 increase more than likely by random chance

Regression Validation

- Recall : Anscombe's quartet
- Need visualization of data to see that the regression has broken down or is not suitable
- Analysis of residuals (visual and numerical)
  - Random or not?
  - Varied with time?
- Additional Reading
  - https://en.wikipedia.org/wiki/Regression_validation

Classification Metrics

- Consider Two Class Problem (yes/no)
  - When true answer is **yes** and your model says **yes**, that is a True Positive
  - When true answer is **no** and your model says **no**, that is a True Negative
  - When true answer is **yes** and your model says **no**, that is a False Negative
  - When true answer is **no** and your model says **yes**, that is a False Positive

Classification Metrics

- Confusion Matrix

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Classification Metrics

- Why a 90% “accurate” model is not always good enough
- My favorite professor’s problem:
  - Land mine detection algorithms for the US Army
  - Yes: It is a land mine
  - No: It is not a land mine

Classification Metrics

- Why a 90% “accurate” model is not always good enough
- My favorite professor’s problem:
  - Land mine detection algorithms for the US Army
  - Yes: It is a land mine
  - No: It is not a land mine
- What is the cost of a False Positive?

Classification Metrics

- Why a 90% “accurate” model is not always good enough
- My favorite professor’s problem:
  - Land mine detection algorithms for the US Army
  - Yes: It is a land mine
  - No: It is not a land mine
- What is the cost of a False Positive?
- What is the cost of a False Negative?

Classification Metrics

- Precision: (PPV) positive prediction value
  - How often is a predicted Yes value correct?
- Recall: (TPR) true positive rate
  - How many of the expected Yes are predicted yes?

n=165	Predicted: NO	Predicted: YES	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Classification Metrics

- F-score (or F_1 Score)
  - Measure of accuracy combining Precision and Recall

$$F_1 = 2 \cdot \frac{\mathrm{PPV} \cdot \mathrm{TPR}}{\mathrm{PPV} + \mathrm{TPR}} = \frac{2\mathrm{TP}}{2\mathrm{TP} + \mathrm{FP} + \mathrm{FN}}$$

n=165	**Predicted: NO**	**Predicted: YES**	
Actual: NO	TN = 50	FP = 10	60
Actual: YES	FN = 5	TP = 100	105
	55	110	

Clustering Metrics

- Cluster Validation
  - https://en.wikipedia.org/wiki/Cluster_analysis#Evaluation_and_assessment
- Distance metric driven
  - Ratios of points to centroids or cluster members
  - Euclidean vs Mahalanobis vs other
- Davies-Bouldin index : average ratio of cluster-to-cluster size versus center distance
- Dunn index: ratio between the minimal inter-cluster distance to maximal intra-cluster distance
- Others

Conclusion

- Measures and analyses of machine learning models are critical before operationalizing
  - Fully understand the model
  - Measure the performances against expected and unexpected data
  - Weigh the consequences of erroneous responses