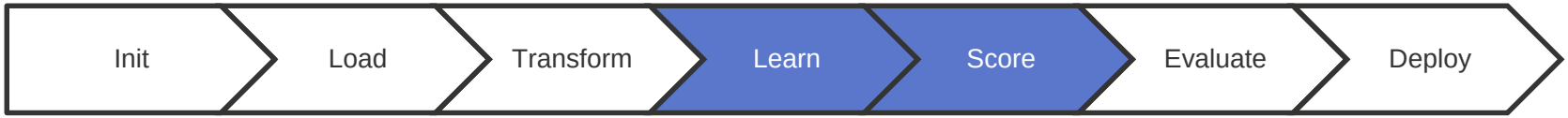
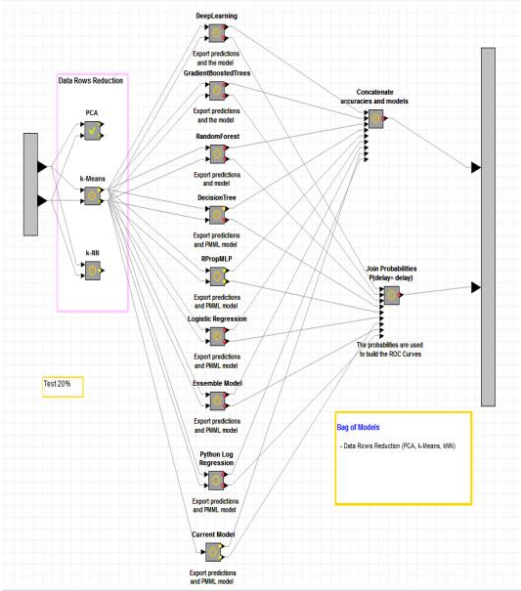
MODEL MANAGEMENT OR „GETTING YOUR LIFE BACK“

Phil Winters
phil.winters@ciagenda.com

COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

Solutions In the Market: Bags of Models

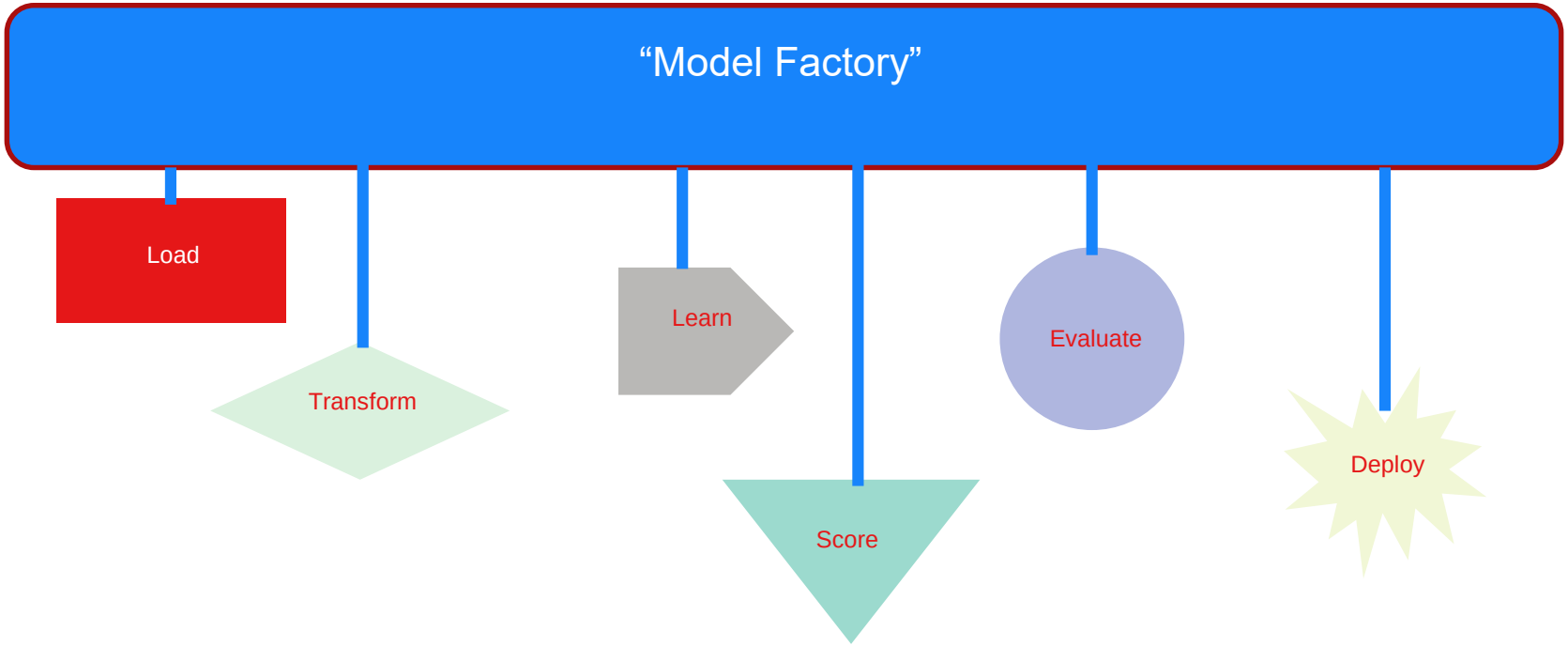




Very Important but just one part of the modelling process!

COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

Solutions In the Market: “Pulling products together”

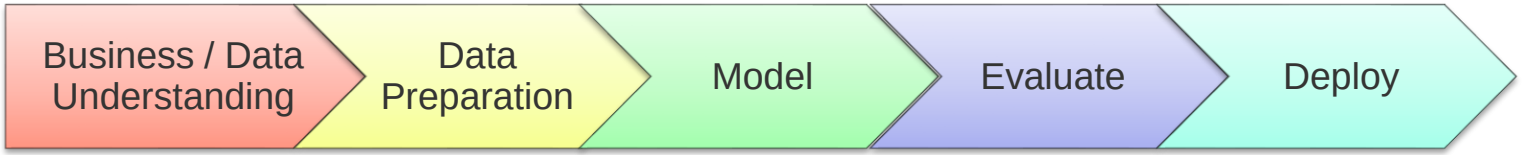


Very Important for some vendors because it corrects major architectural design flaws

COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

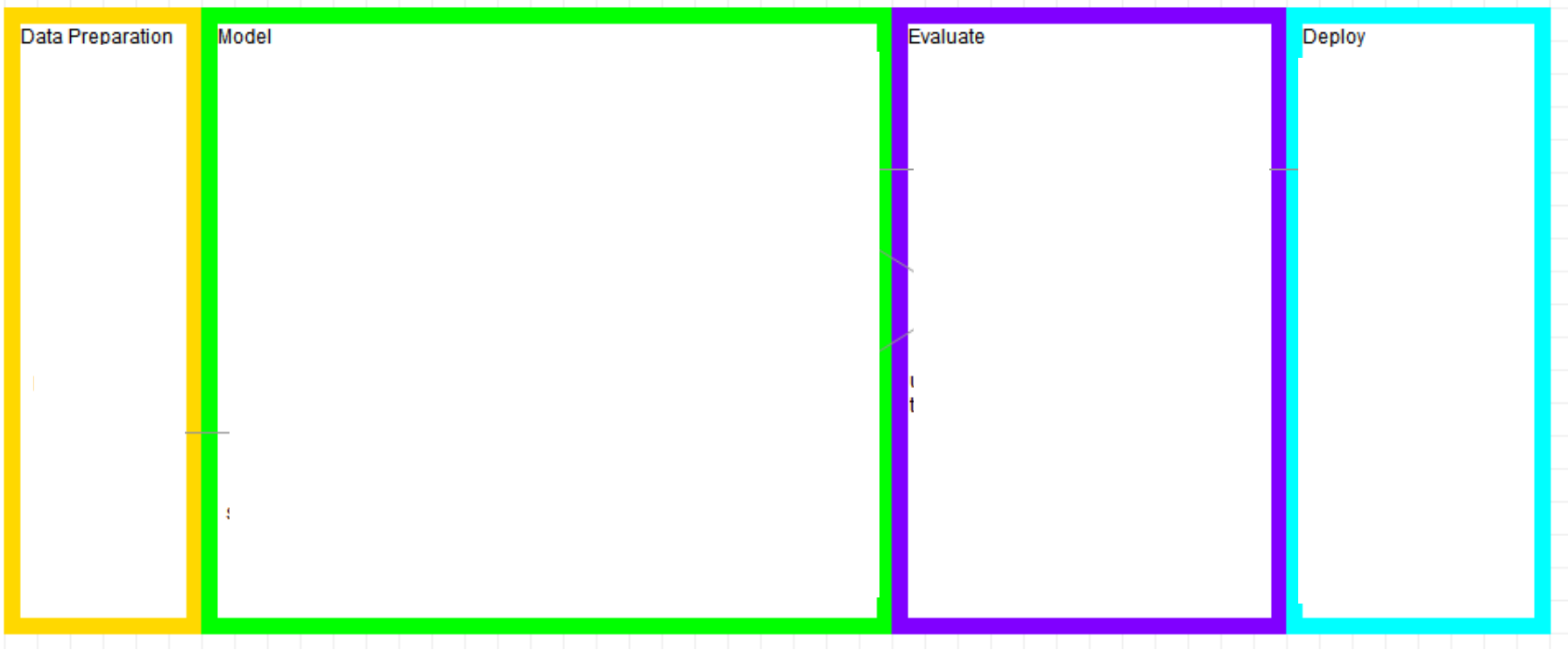
Predictive History: The Analytic Process Wars…

- KDD-Process
- SEMMA
- CRISP ✔



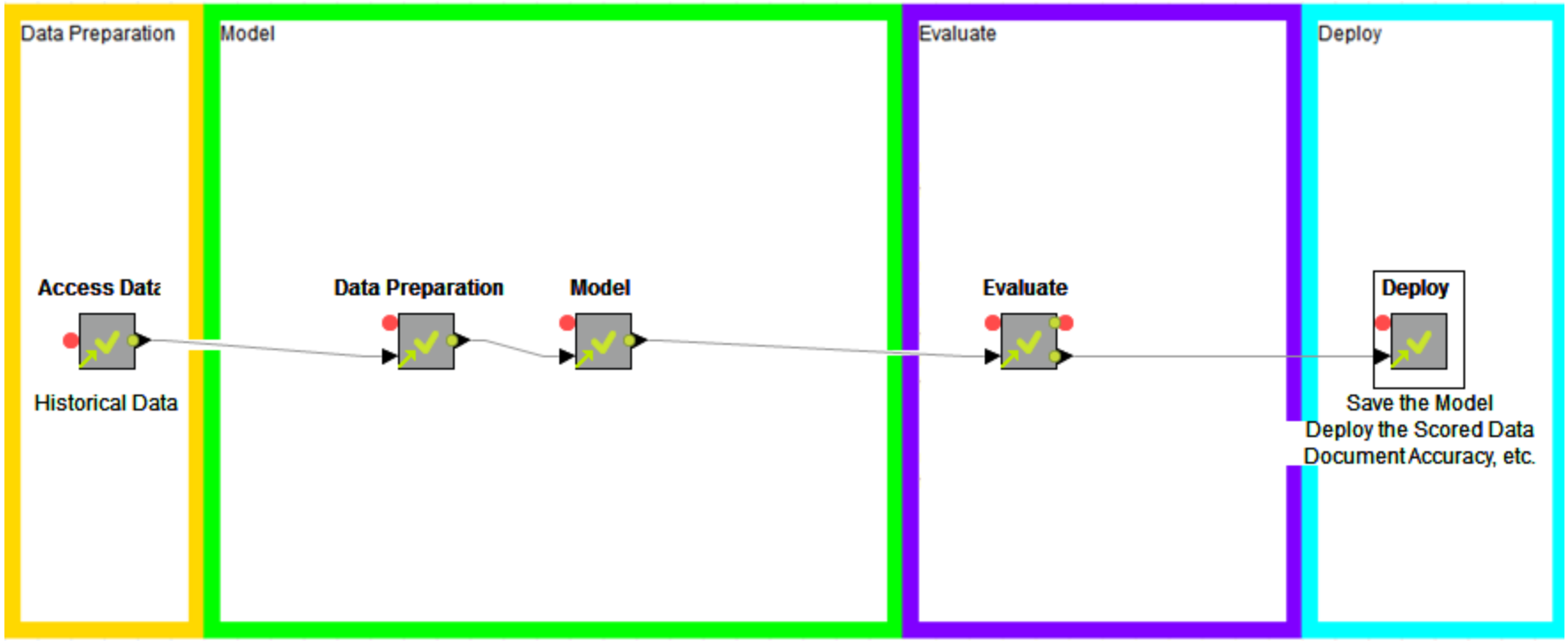
COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

CRISP with your favorite package



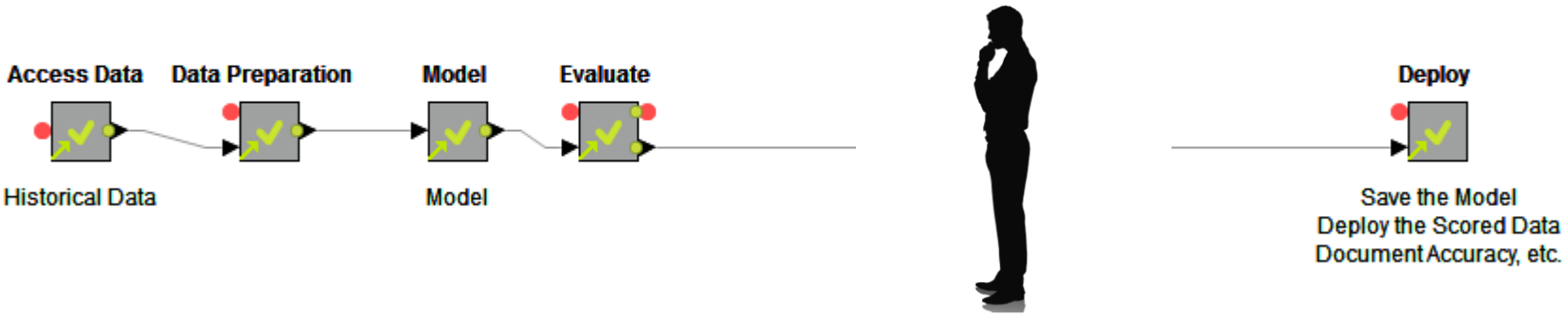
COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

CRISP with your favorite package



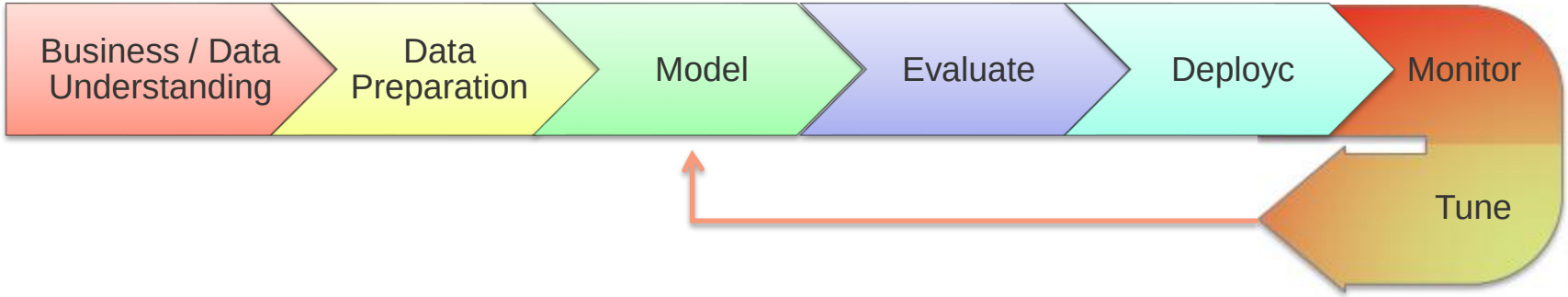
COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

CRISP with your favorite package

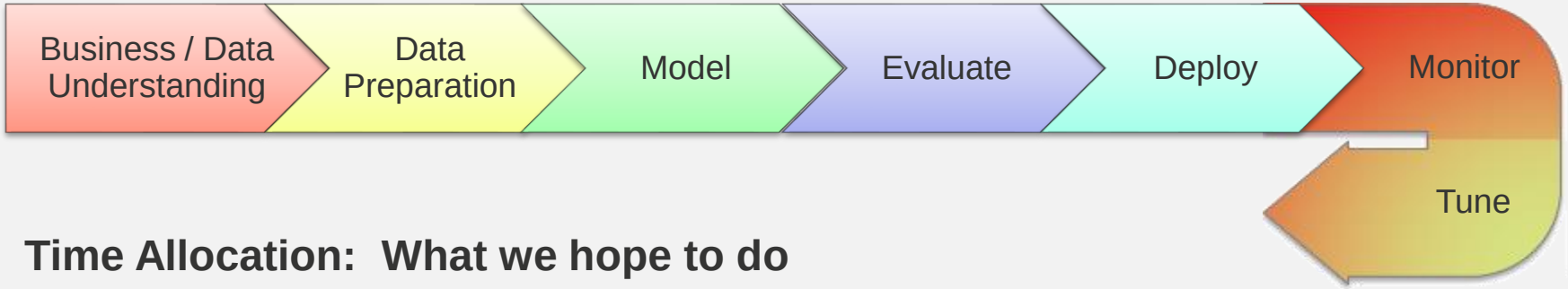


COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

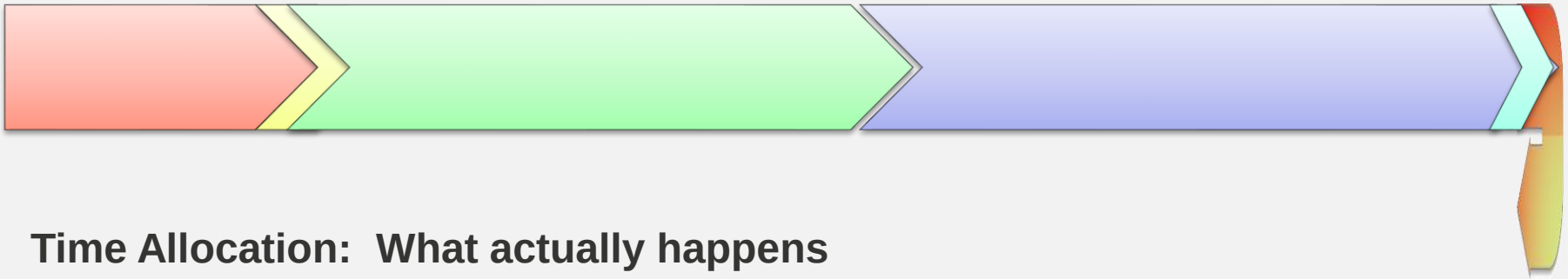
CRISP and Reality: Something missing…



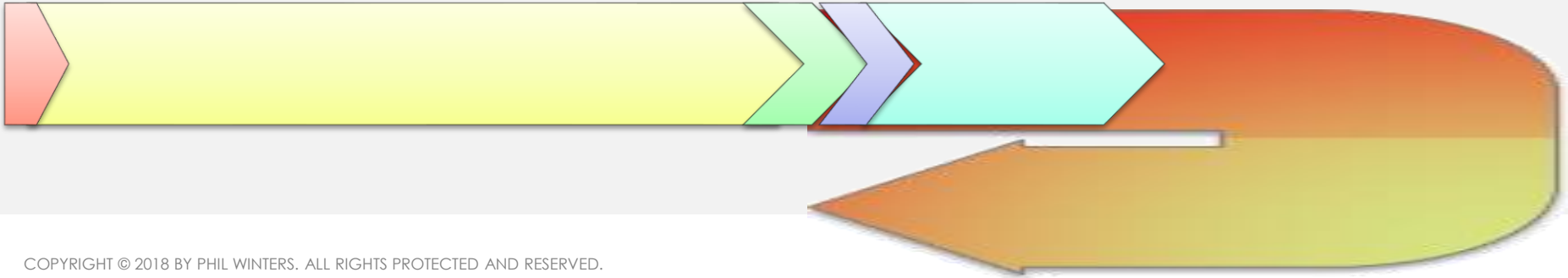
COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.



Time Allocation: What we hope to do

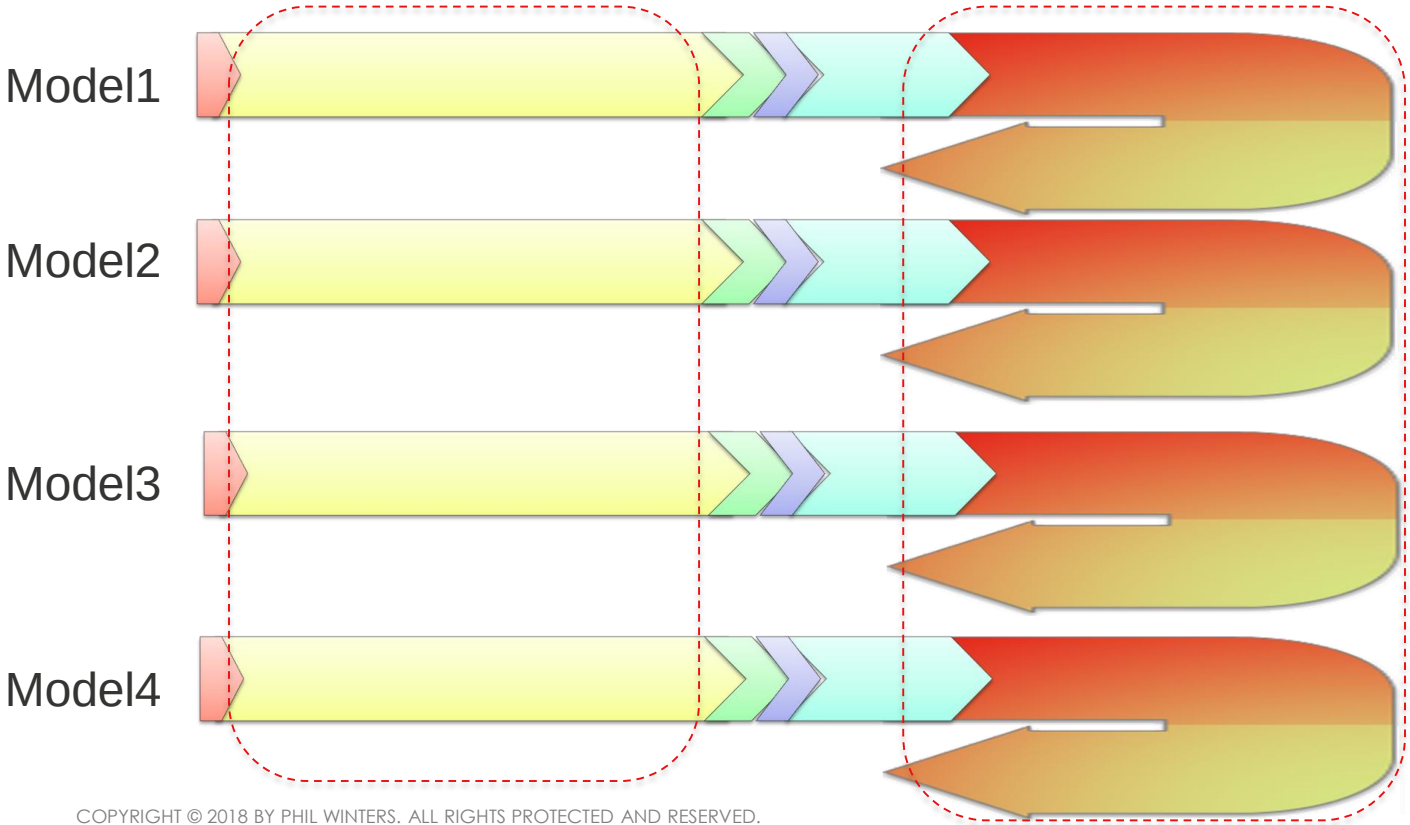


Time Allocation: What actually happens



COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

Your Life = 4 models??????



80% OF YOUR WORK LIFE DOING THE THINGS THAT YOU HATE...

COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

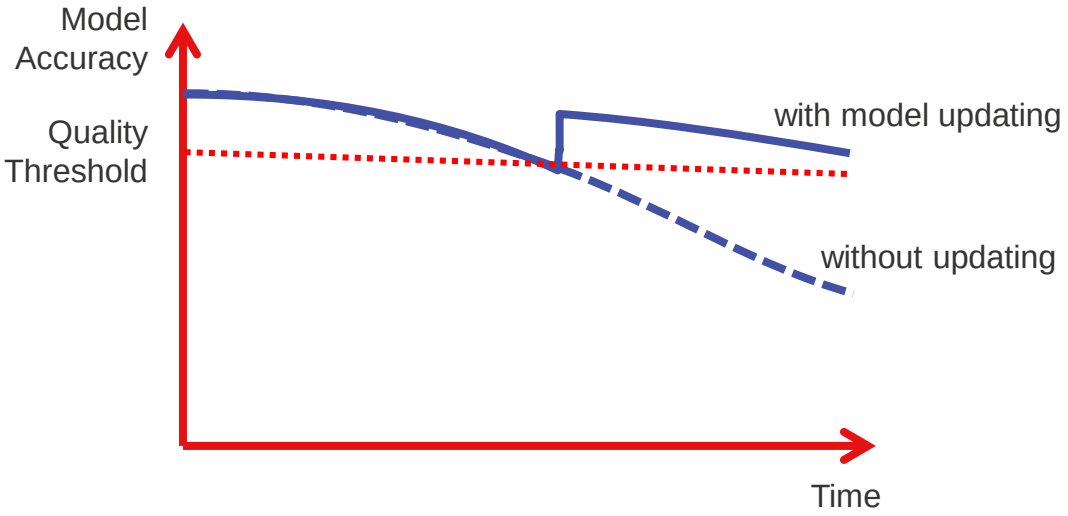
Models

...capture a snap shot of the data (today's reality)

...need to be updated when data (reality) changes

COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

Models in Time



COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

Model Updating

Issues:

- Training of new Model
  vs Updating of Existing One
- Training/Updating/Replacing every time?
- Evaluation on Old
  vs. New vs. Benchmark Data

COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

Models

and it gets worse:

...models rely on specific data preparations...

...deployment is an entire drama in itself. (offline vs. real-time scoring, (web)service, executable, model standards, scalability...)

COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

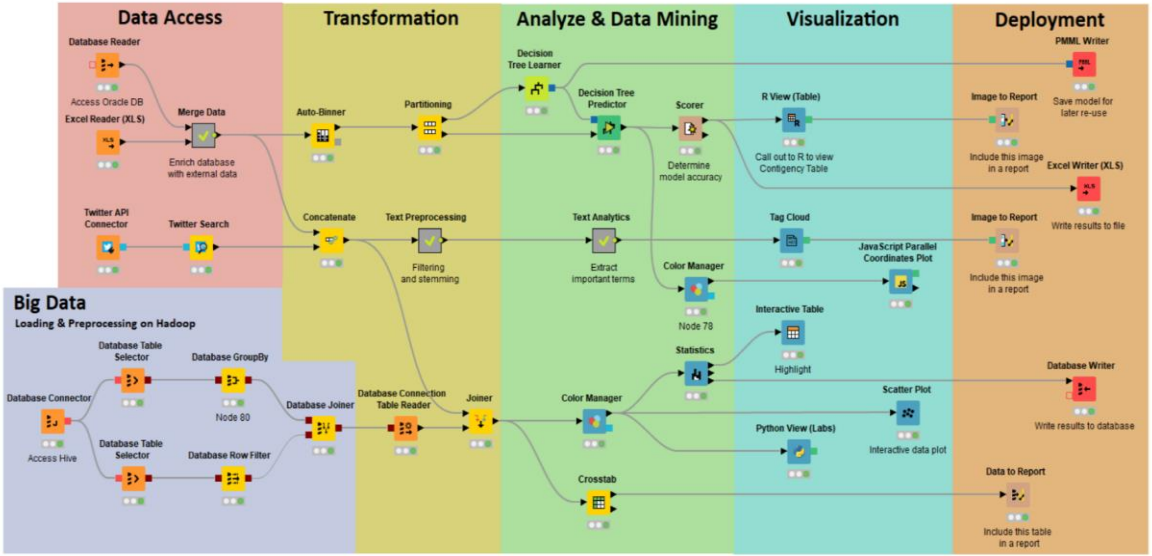
My Life: Easy Peazy Lemon Squeazy….

- Advanced Price Prediction (Forecasting)
- 29000 models
- Built, Trained, Scored, Tested, Monitored, Retrained

- Run Daily in 45 Minutes
- Totally Automated “Hands Off”

COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

Model Learning: a concrete Example



COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

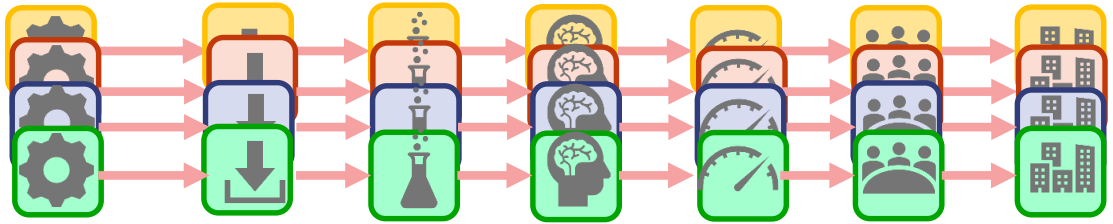
More Model Processes

Tough choice:

- one pipeline fits all

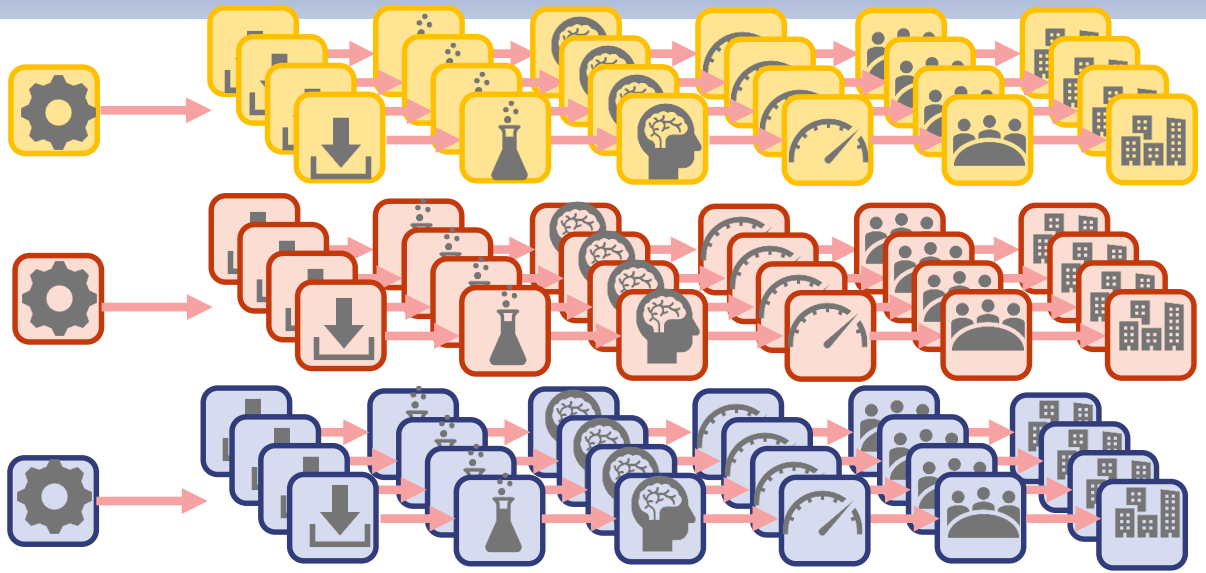
vs.

- individual pipelines for all.



COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

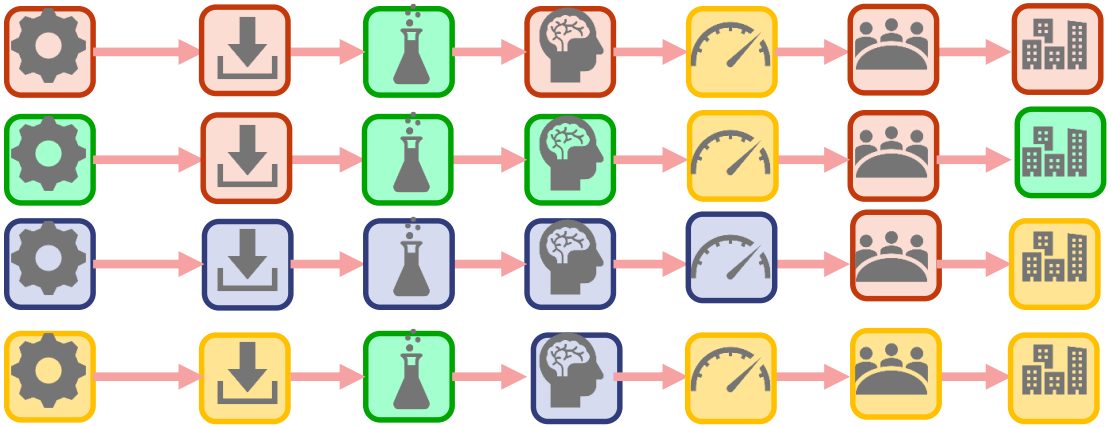
Model Families



- Issues with many families:
  too many pipelines with likely repetitive pieces...

COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

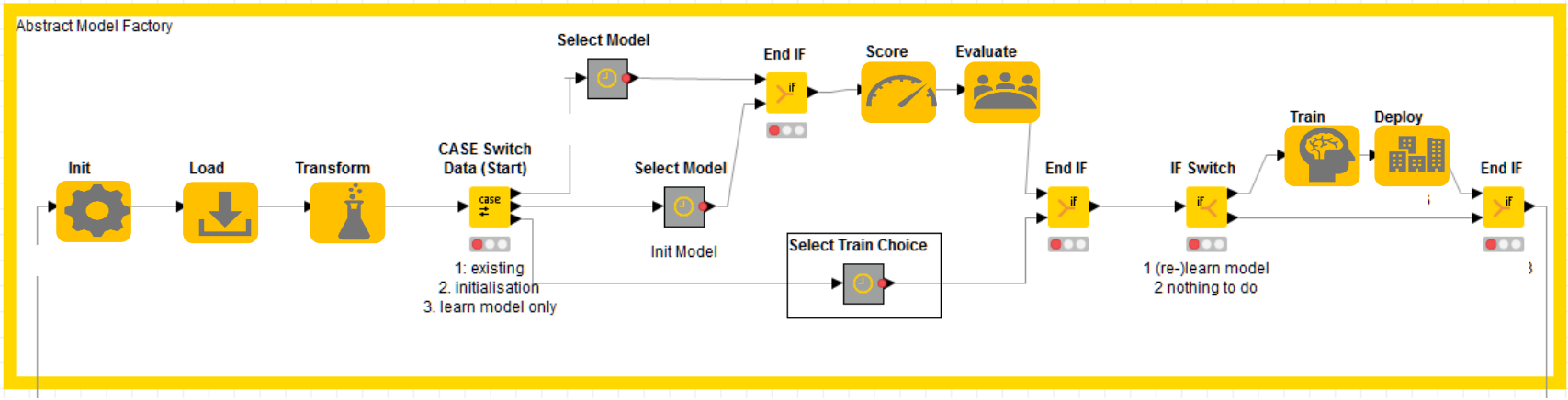
Model Factories



- Pipeline = sum of its pieces.
- Reuse what's reusable.

COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

Model Management Process



COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

Model Factory Concepts + Terminology

Process Definition:

Process 1 (“Purchase Propensity”)
Process 2 (“...”)
Process 3 (“Churn Pred.”)

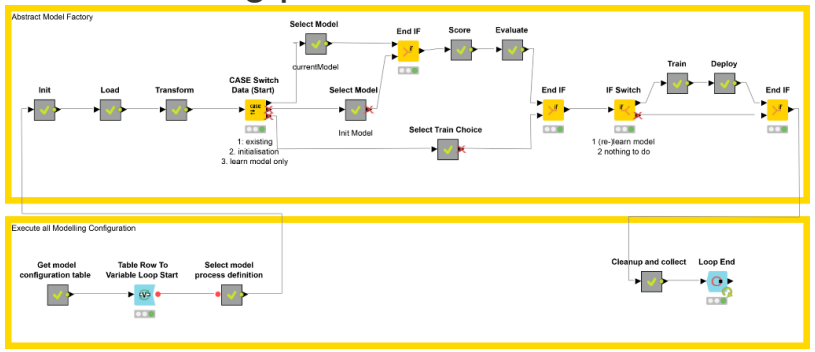
Modeling Process Configuration:

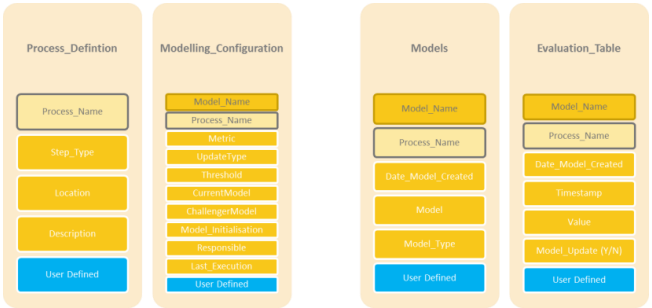
Product1	Process1 (parameters...)
Product2	Process1 (parameters...)
Product3	Process1 (parameters...)
Topic4	Process2 (parameters...)
Topic5	Process2 (parameters...)
New1	Process 3(parameters…)
….	

COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

The Next Generation Model Factory

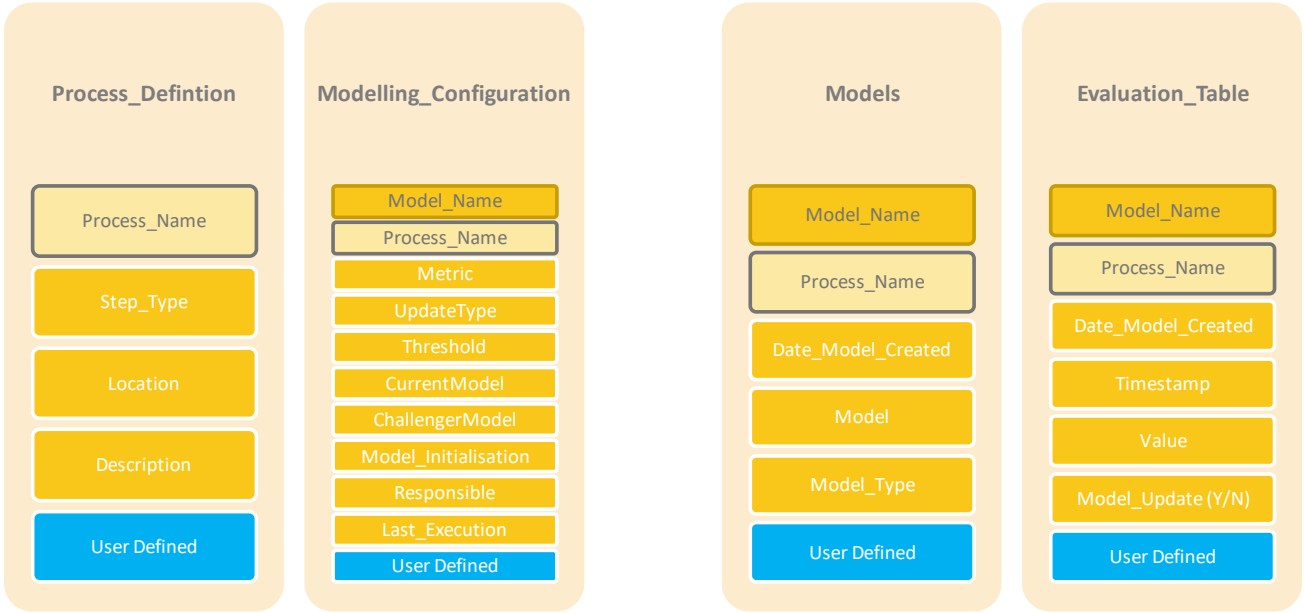
Using a platform and well defined metadata to create, execute, monitor and control all aspects of modelling processes





COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

The Tables



COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

Model_Process Table

Process 1 (“Customer Seg”)
Process 2 (“…”)
Process 3 (“Churn Pred.”)

Process_Name	Step_Type	Location
AirBNB RentApp Prediction	INIT	/AirBNB/workflows/rentedAppPred/AirBnB_rentAppPred_01_Init
AirBNB RentApp Prediction	LOAD	/AirBNB/workflows/rentedAppPred/AirBnB_rentAppPred_02_Load
AirBNB RentApp Prediction	TRANSFORM	/AirBNB/workflows/rentedAppPred/AirBnB_rentAppPred_03_ET
AirBNB RentApp Prediction	LEARN	/AirBNB/workflows/rentedAppPred/AirBnB_rentAppPred_05_Training
AirBNB RentApp Prediction	SCORE	/AirBNB/workflows/rentedAppPred/AirBnB_rentAppPred_04_Evaluation
AirBNB RentApp Prediction	EVALUATE	/_Process_Step_Templates/workflows/Templates/06_Evaluate_take_firstRowAndCell
AirBNB Price Prediction Cities	INIT	/AirBNB_Cities/workflows/Predict_App_Price/01_Init_for_Cities
AirBNB Price Prediction Cities	LOAD	/AirBNB_Cities/workflows/Predict_App_Price/02_Load_for_Cities
AirBNB Price Prediction Cities	TRANSFORM	/AirBNB/workflows/PricePred/AirBnB_PricePred_03_ET
AirBNB Price Prediction Cities	LEARN	/AirBNB/workflows/PricePred/AirBnB_PricePred_05_Training
AirBNB Price Prediction Cities	SCORE	/AirBNB/workflows/PricePred/AirBnB_PricePred_04_Evaluation
AirBNB Price Prediction Cities	EVALUATE	/_Process_Step_Templates/workflows/Templates/06_Evaluate_take_firstRowAndCell

COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

Model_Configuration Table

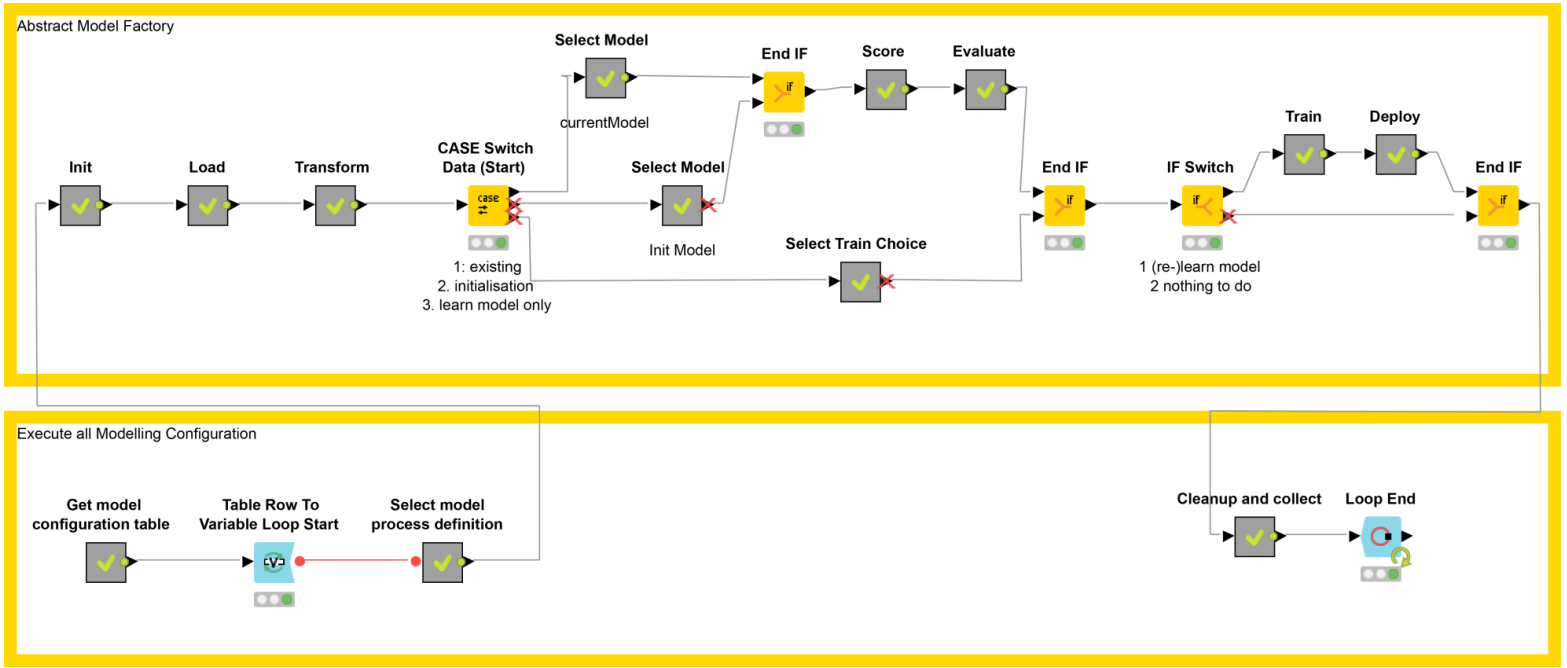
Modeling Process Configuration:

Product1 Process1 (parameters...)
Product2 Process1 (parameters...)
Product3 Process1 (parameters...)
Topic4 Process2 (parameters...)
Topic5 Process2 (parameters...)
New1 Process 3(parameters...)
....

Model_Name	Process_Name	Metric	Threshold	Current...	Update_Type	Model_Initialisation	ChallengerModel	Responsible	Last_Execution
Empty_Template	_Template	none	0.667	?	?	?	?	iris.adae@knime.com	?
Hong_Kong	AirBNB Price Prediction Cities	RMSD	16.172	?	Automatic	all	all	modelFactory@knime.com	2017-03-12;00:40:39.473
Seattle	AirBNB Price Prediction Cities	RMSD	1.13	?	Automatic	all	all	modelFactory@knime.com	2017-03-12;00:41:07.675
Vancouver	AirBNB Price Prediction Cities	RMSD	0.36	?	Automatic	all	all	modelFactory@knime.com	2017-03-12;00:41:52.390
Venice	AirBNB Price Prediction Cities	RMSD	0.259	?	Automatic	all	all	modelFactory@knime.com	2017-03-12;00:42:30.529
all	AirBNB Price Prediction Cities	RMSD	3.176	?	Automatic	?	?	?	2017-03-12;00:42:49.153
2_Beds	AirBNB Price Prediction	RMSD	0.538	?	Automatic	?	?	iris.adae@knime.com	2017-03-12;00:43:22.296
>2_Beds	AirBNB Price Prediction	RMSD	2.414	?	Automatic	?	?	iris.adae@knime.com	2017-03-12;00:43:39.25
CS_TV_000127	Product Price Trend	RMSD	0	?	Automatic	CS_TV_002128	?	phil.winters@knime.com	2017-03-12;00:44:06.966
CS_TV_002128	Product Price Trend	RMSD	0	?	Automatic	CS_TV_000127	?	phil.winters@knime.com	2017-03-12;00:44:24.550
IE_WM_002651	Product Price Trend	RMSD	0	?	Automatic	PC_WM_001953	?	phil.winters@knime.com	2017-03-12;00:44:49.73
PC_WM_010138	Product Price Trend	RMSD	1	?	Automatic	PC_WM_001953	?	phil.winters@knime.com	2017-03-12;00:45:22.492
>2_Beds	AirBNB RentApp Prediction	RMSD	26.136	?	Automatic	?	?	iris.adae@knime.com	2017-03-12;00:46:12.738
2_Beds	AirBNB RentApp Prediction	RMSD	98.853	?	Automatic	?	?	iris.adae@knime.com	2017-03-12;00:46:24.704
Boston	AirBNB Price Prediction Cities	RMSD	3.565	?	Automatic	all	?	?	2017-03-12;00:47:21.674
Amsterdam	AirBNB Price Prediction Cities	RMSD	8.164	?	Automatic	all	all	modelFactory@knime.com	2017-03-12;00:48:11.664
Geneva	AirBNB Price Prediction Cities	RMSD	0.475	?	Automatic	all	all	modelFactory@knime.com	2017-03-12;00:48:40.8
IE_TV_008995	Product Price Trend	Trend	1	?	Manual	?	?	test@knime.com	2017-03-12;00:49:13.709
PC_WM_001953	Product Price Trend	RMSD	1	?	Automatic	PC_WM_010138	?	phil.winters@knime.com	2017-03-12;00:49:29.256

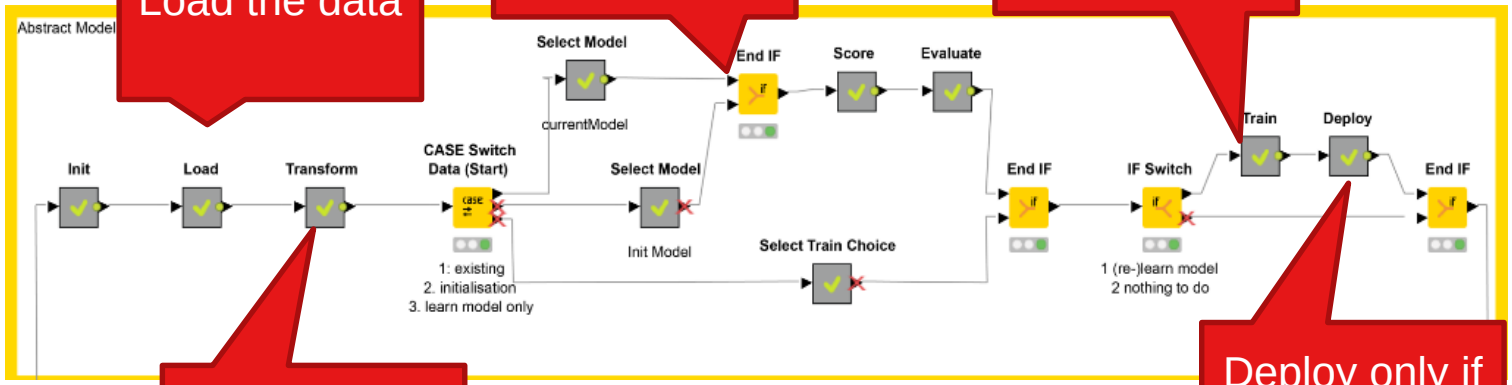
COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

The KNIME Model Factory



COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

The Model Factory



Load the data

Evaluate only if model exists

Retrain if error exceeds threshold

Split into training and test set

Deploy only if update type allows it

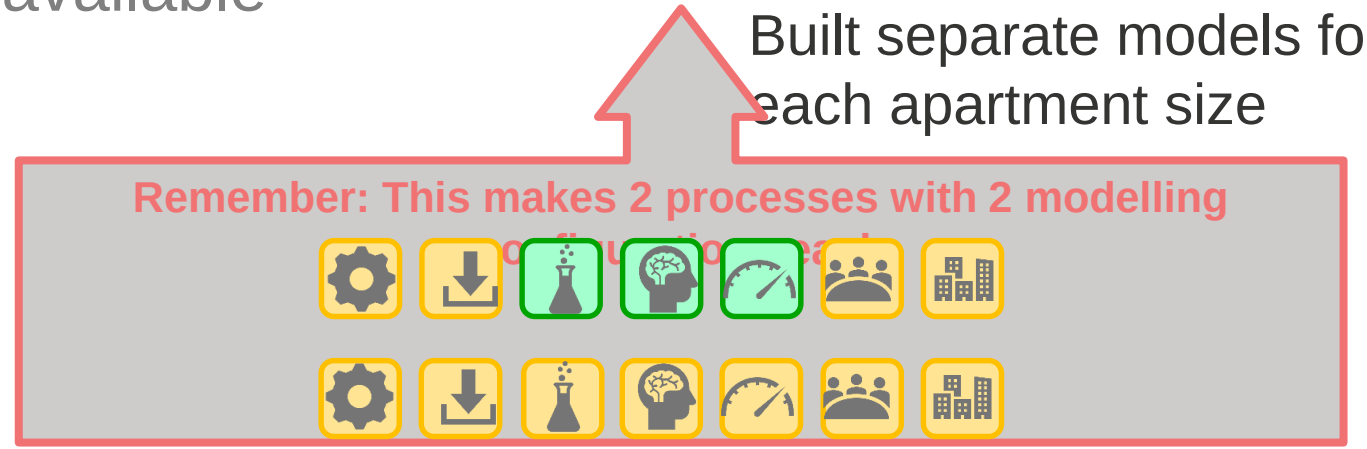
COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

Use Case: AirBnB

- Data is freely available

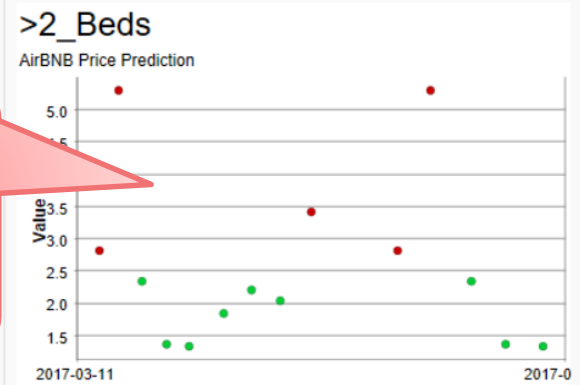
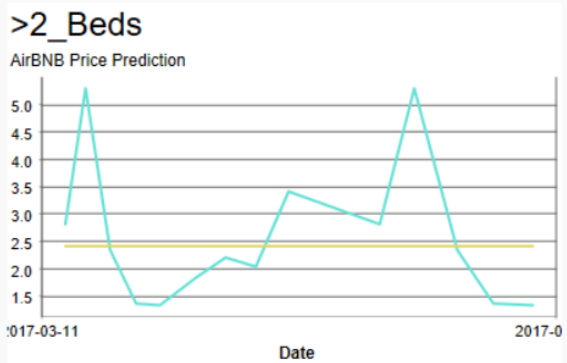
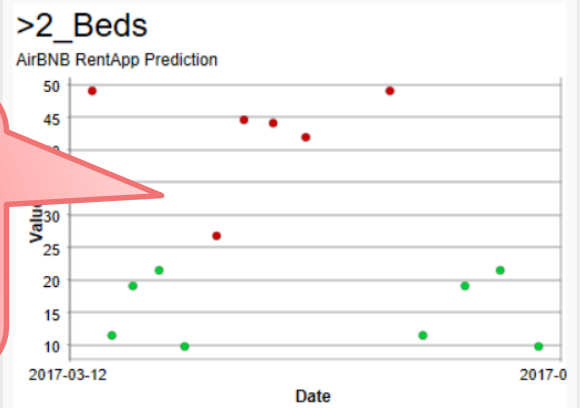
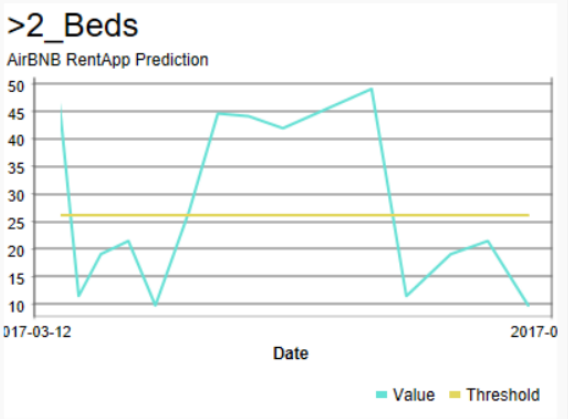
Tasks

- Predict the appartment price
- Predict availabilty
- Built separate models for each apartment size



COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

Results - AirBNB

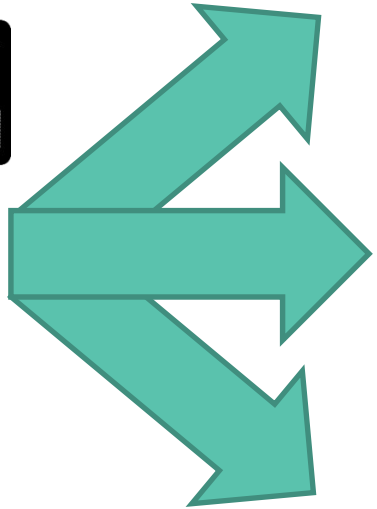
>2_Beds	AirBNB Price Prediction		
			
>2_Beds	AirBNB RentApp Prediction		

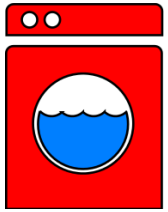
Accuracy for predicting with deployed model

Each red datapoint Resulted in a retrain of the model

COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

Use case: Trend prediction





- Price-Trend prediction
  - Predict if the price will go up, down or stay

Built one model per product

Retrain and deploy the model dynamically

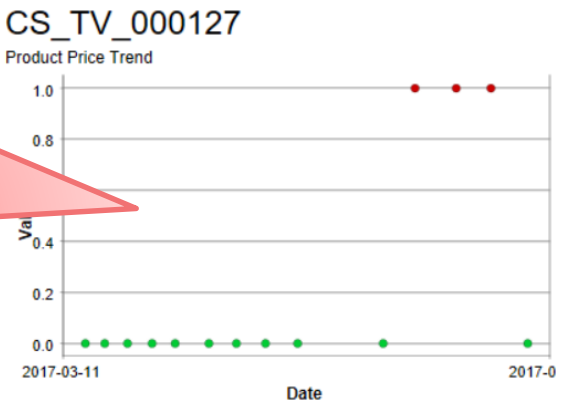
COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

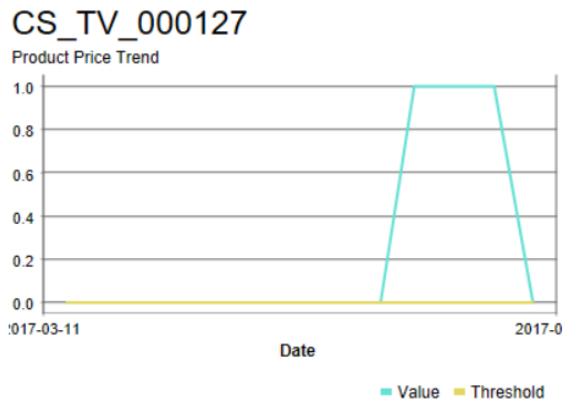
Results – Product Data

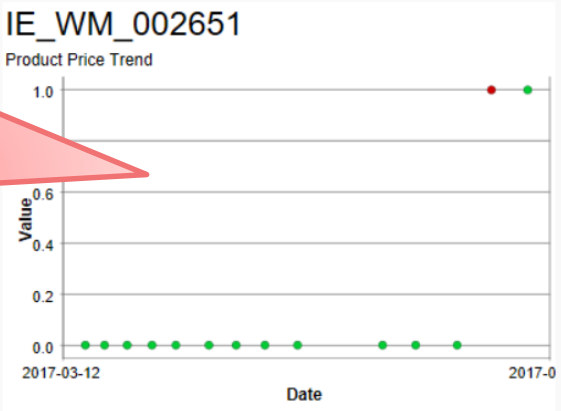
CS_TV_000127	Product Price Trend
IE_WM_002651	Product Price Trend

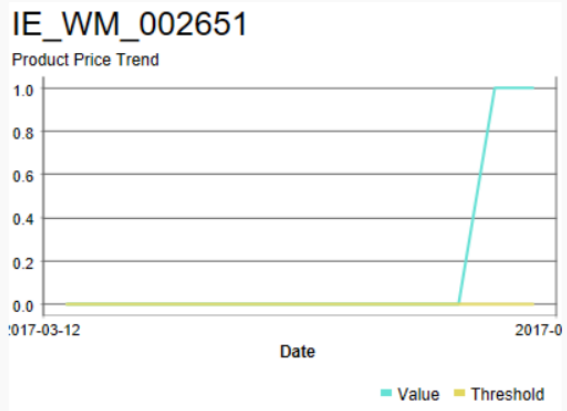
Accuracy for predicting with deployed model

Each red datapoint Resulted in a retrain of the model









COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

Modelling Process on different products

Same modelling workflow applied to multiple items.

- One Model Process
  - same workflow
  - Init configured to read the Configuration_Name
- Multiple Model Configurations
  - One for each Product or item
  - Configuration_Name = Product Name
  - Many products? Have a workflow build the model Configurations!

COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

A model MUST be in production for a new product

- A score must appear in production for a new item, but there is not enough data to create a model score.

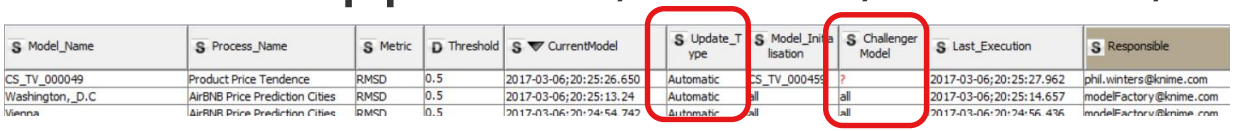
Model_Name	Process_Name	Metric	Threshold	CurrentModel	Update_Type	Model_Initialisation	Challenger Model	Last_Execution	Responsible
CS_TV_000049	Product Price Tendence	RMSD	0.5	2017-03-06;20:25:26.650	Automatic	CS_TV_000459		2017-03-06;20:25:27.962	phil.winters@knime.com
Washington,_D.C	AirBNB Price Prediction Cities	RMSD	0.5	2017-03-06;20:25:13.24	Automatic	all		2017-03-06;20:25:14.657	modelFactory@knime.com
Vienna	AirBNB Price Prediction Cities	RMSD	0.5	2017-03-06;20:24:54.742	Automatic	all		2017-03-06;20:24:56.436	modelFactory@knime.com

- Process Configuration
  - Specify an existing Model_Initialization Process
  - This chooses a “similar” process configuration
  - Runs on the alternate data until there is enough to “retrain” on new data
    - Switch happens automatically.
  - More sophisticated might be for INIT to do a similarity search, then use the “closest” process configuration

COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

Challenge an existing model process

- with a new approach, model, method, data etc.



Model_Name	Process_Name	Metric	Threshold	CurrentModel	Update_Type	Model_Initialisation	Challenger Model	Last_Execution	Responsible
CS_TV_000049	Product Price Tendence	RMSD	0.5	2017-03-06;20:25:26.650	Automatic	CS_TV_000459	?	2017-03-06;20:25:27.962	phil.winters@knime.com
Washington, _D.C	AirBNB Price Prediction Cities	RMSD	0.5	2017-03-06;20:25:13.24	Automatic	all	all	2017-03-06;20:25:14.657	modelFactory@knime.com
Vienna	AirBNB Price Prediction Cities	RMSD	0.5	2017-03-06;20:24:54.742	Automatic	all	all	2017-03-06;20:24:56.436	modelFactory@knime.com

  - Create a new Model Process
  - Create a new Model Configuration
    - Set Update_Type = Challenger
    - A Challenger will never be Deployed !
  - Point the Champion Configuration to the Challenger
    - Statistics will be captured for both
  - When you are happy, swap them out.

COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

Manually Retrain process

- A production model_configuration that is only retrained when triggered

Model_Name	Process_Name	Metric	Threshold	CurrentModel	Update_Type	Model_Initialisation	Challenger Model	Last_Execution	Responsible
CS_TV_000049	Product Price Tendence	RMSD	0.5	2017-03-06;20:25:26.650	Automatic	CS_TV_000459	?	2017-03-06;20:25:27.962	phil.winters@knime.com
Washington,_D.C	AirBNB Price Prediction Cities	RMSD	0.5	2017-03-06;20:25:13.24	Automatic	all	all	2017-03-06;20:25:14.657	modelFactory@knime.com
Vienna	AirBNB Price Prediction Cities	RMSD	0.5	2017-03-06;20:24:54.742	Automatic	all	all	2017-03-06;20:24:56.436	modelFactory@knime.com

  - Model Process Configuration
    - Update_Type is set to MANUAL
      - Because we want to continue scoring and capturing information to evaluate
      - No Retraining will occur
    - Reset Type to MANUAL_RETRAIN
      - Retraining will then occur
      - At the end of the model process, the type will be reset to MANUAL

COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

Eliminating/Extending a process step from a Process Model

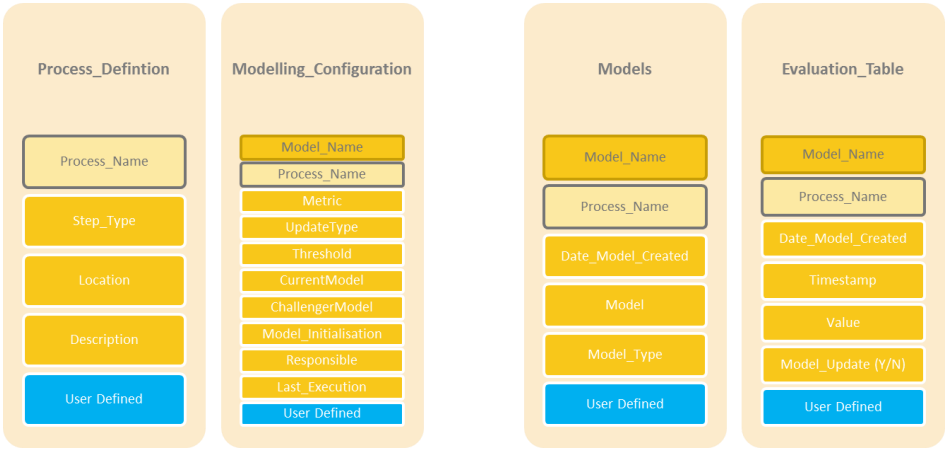
- Because the process is fundamentally different

- Create a new Model Process
  - Use “Empty” step templates to eliminate steps
  - Alternately, extend the Abstract Modelling Workflow

COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

More Requirements for the Modelling Factory

- We have security, standards, teams, requirements…
- Extend the appropriate table with User-Defined Fields
  - These fields are then available within your model factory all as flow variables



COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

My Life: Easy Peazy Lemon Squeazy….

- Advanced Price Prediction (Forecasting)
- 29000 models
- Built, Trained, Scored, Tested, Monitored, Retrained

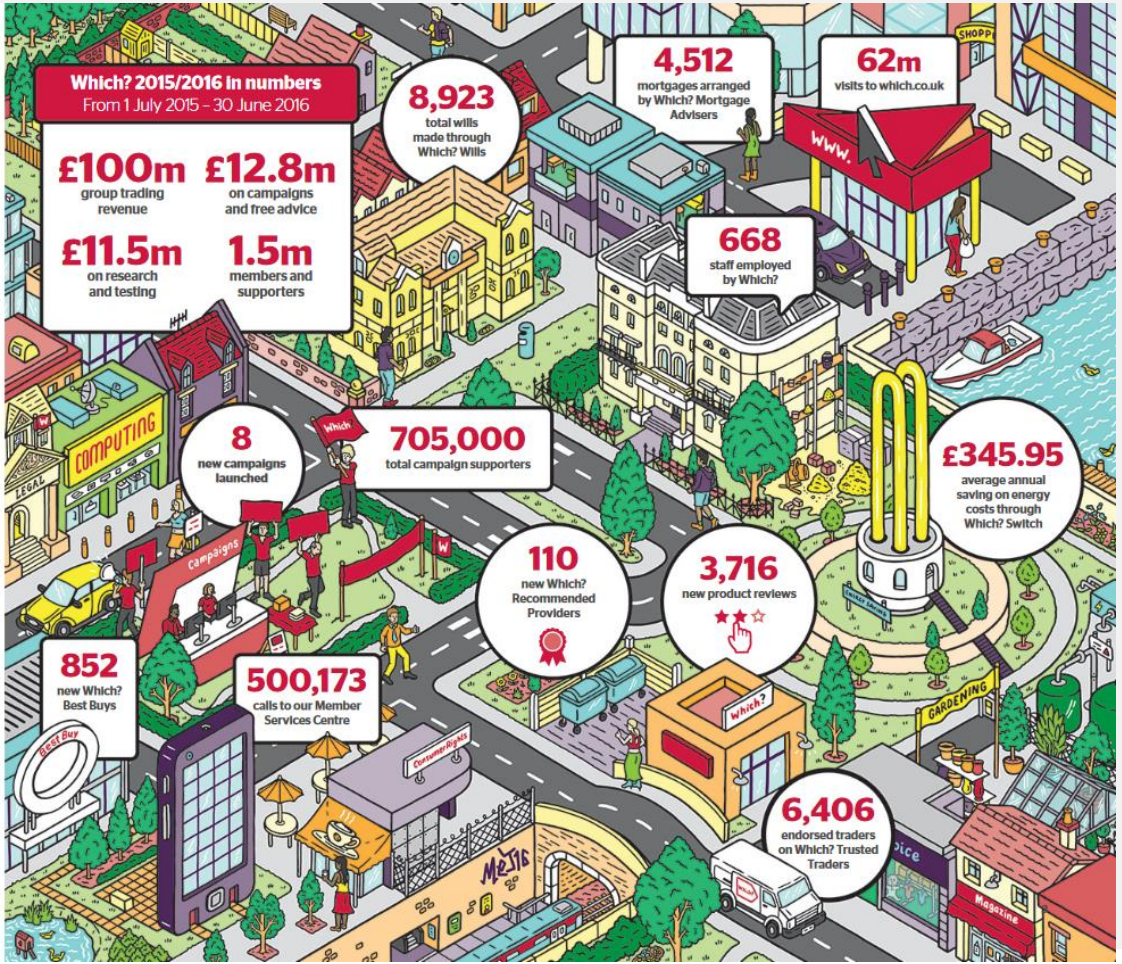
- Run Daily in 45 Minutes
- Totally Automated “Hands Off”

COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

My Life with Which? Consumer Association UK

COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

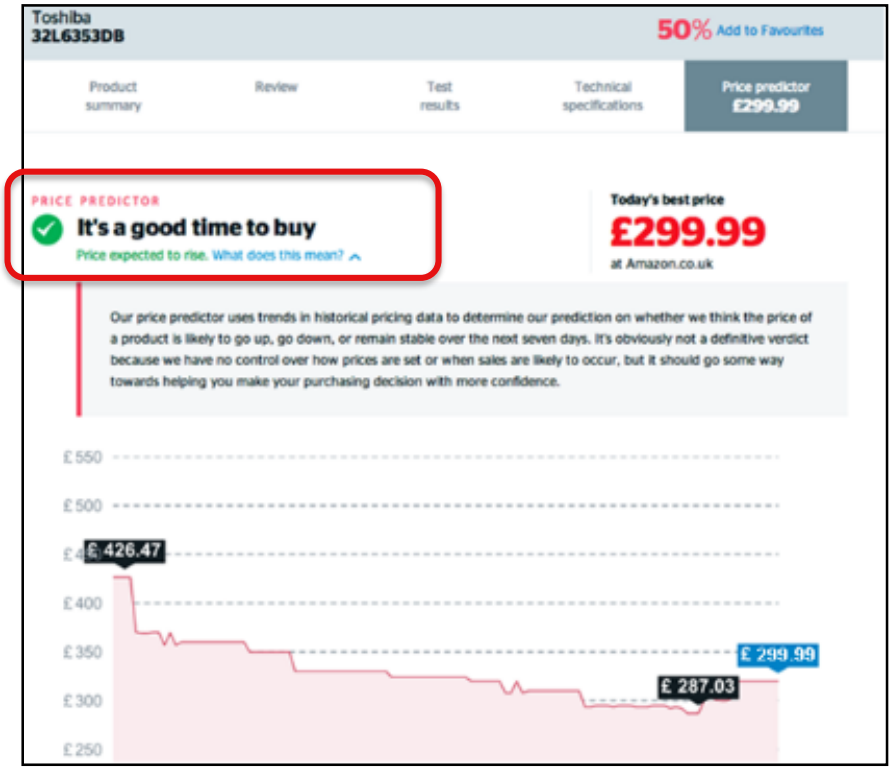
Which? exists to make individuals as powerful as the organisations they deal with in their daily lives



Which?

COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

Ensemble of Decision Tree + Linear Regression



"Which? members tell us the price prediction is the most value service on the website. "

COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

Model Factory / Management / Automation Options

- Build your own system
  - With your favorite Package
- Take an existing, free, open source example
  - KNIME Analytics Platform
- Buy a Package
  - KNIME Server (to schedule the factory)
  - IBM SPSS Deployment Manager on IBM SPSS
  - SAS Model Management on SAS
  - Fico (but its all magic…)
  - Others? Talk to me !!!!!

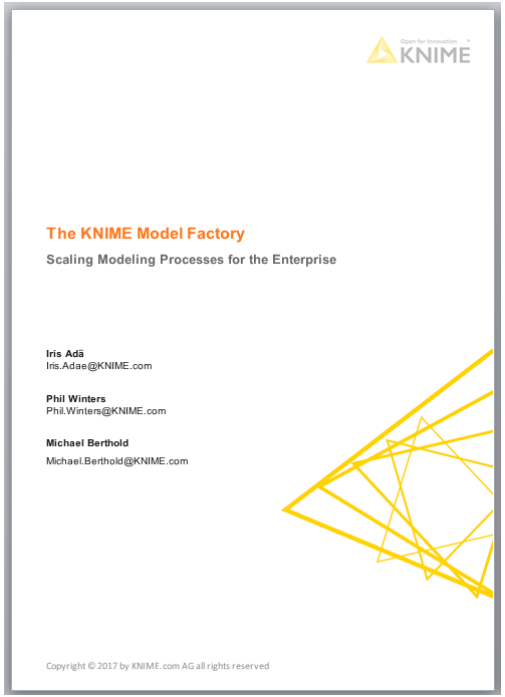
COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

What's available: Model Factory

Examples:

- 26_Model_Process_Management
  - Factory
    - _Configuration_Template
    - Model_Factory
  - helperWorkflows
    - tempDataCopies
    - SetThresholdsToMedian
    - Webportal_CheckState
    - Webportal_SeeAll
    - Webportal_Update
    - Webportal_Update_Type_Change
  - Metainfo
    - Create_Log_Tables
    - Edit_Tables_Locally
    - Evaluation_Table.table
    - Modelling_Configuration.table
    - Models.table
    - Process_Definition.table
  - Process
    - _Metanodes
    - _Process_Step_Templates
    - AirBNB
    - AirBNB_Cities
    - Products

White paper



Next step:
Download it, learn and customize it for your own needs!

www.knime.com/whitepapers

COPYRIGHT © 2018 BY PHIL WINTERS. ALL RIGHTS PROTECTED AND RESERVED.

phil.winters@ciagenda.com