FirstEnergy Corp.
Call Volume Forecast
Using RapidMiner

Michael Stansky   ~~**Brian Riordan**~~
Consultant, Data Analytics   ~~**Advisor, Corporate Risk**~~
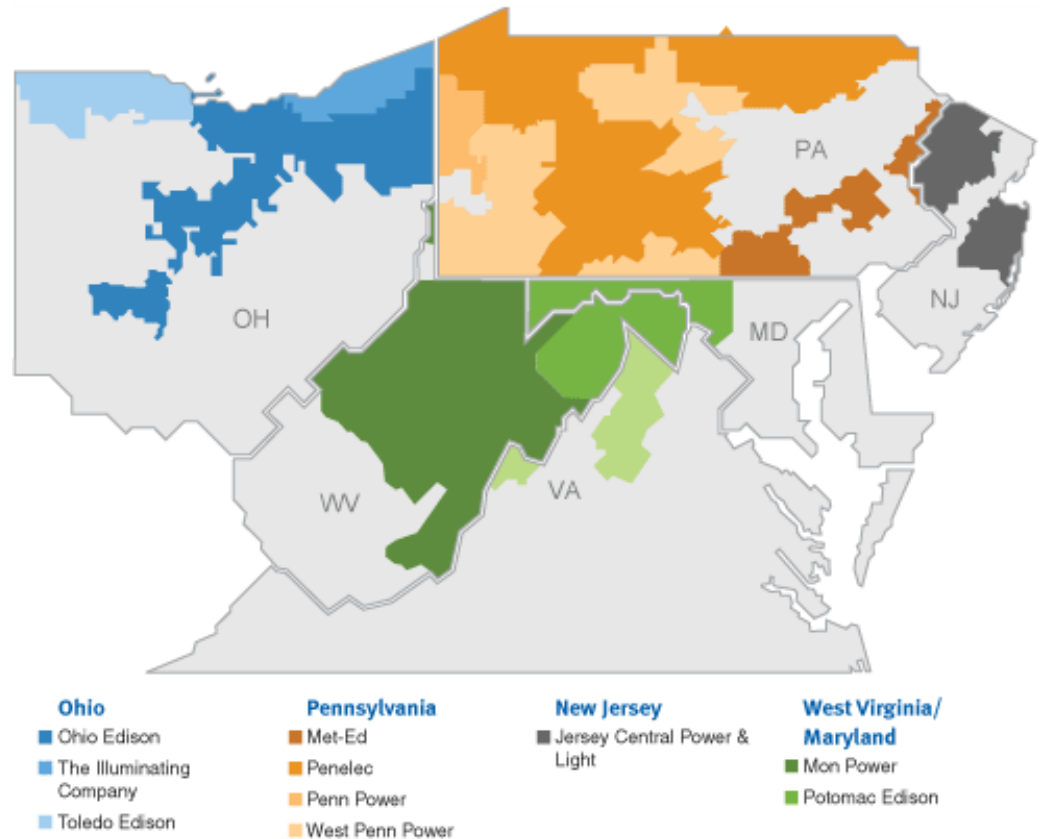
February 12, 2020

Contents

- **FirstEnergy Introduction**

- **History and use of RapidMiner at FirstEnergy**

- **Our First Enterprise Project: Call Volume Forecast**

- **Engagement with Clarkston Consulting**

- **Moving Forward**

FirstEnergy at a Glance

FirstEnergy's 10 regulated companies form one of the nation's largest investor-owned electric systems serving customers in the Midwest and Mid-Atlantic regions. Stretching from western Ohio to the New Jersey shore, and from northern Pennsylvania to southern West Virginia, the companies operate a vast infrastructure of more than 269,000 miles of electric distribution lines and are dedicated to providing our 6 million customers with safe, reliable and responsive service.

Ohio
- Ohio Edison
- The Illuminating Company
- Toledo Edison

Pennsylvania
- Met-Ed
- Penelec
- Penn Power
- West Penn Power

New Jersey
- Jersey Central Power & Light

West Virginia/ Maryland
- Mon Power
- Potomac Edison

OUR MISSION

We are a forward-thinking electric utility powered by a diverse team of employees committed to making customers' lives brighter, the environment better and our communities stronger.

RapidMiner & Analytics Timeline

- **Pre-2017**
  - SAS, R, Matlab, Minitab and others were used in pockets across the organization
  - The Quant Group, a cross-functional organization of analytic employees begins meeting

- **2017**
  - Began evaluating a single, user-friendly solution for advanced analytics
  - Conducted an RFP and use case for a visual analytics platform
  - Reviewed several tools and selected RapidMiner in late 2017

- **2018:**
  - Began rollout and training to small group of 10-20 users
  - Ad hoc work with RapidMiner Studio
  - Selected "Call Volume Forecasting" as a use case to showcase value
  - Began hosting a monthly RapidMiner User Group in December 2018

- **2019:**
  - Began hosting bi-annual Hack-a-thon's where RapidMiner is one of several tools
  - Reached ~50 RapidMiner users
  - Two fully enterprise production processes, several in staging and many in development

Uses of RapidMiner Across FirstEnergy

- **Long-term Planning and Analytics**

- **Regulated Commodity Sourcing**

- **Asset Information and Reporting**

- **Revenue Operations**

- **Corporate & Enterprise Risk**

- **Smart Meter Operations / Meter Reporting**

- **Rates and Regulatory Affairs/Settlements**

- **Transmission and Physical Security**

- **Customer Service & Contact Center Analytics**

Customer Contact Center Operations

- **FirstEnergy employs ~500 full-time and ~200 contractors in its three customer contact centers located in Ohio, Pennsylvania and West Virginia.**

- **In a typical year, FirstEnergy's customer contact centers handle over 16 million customer calls.**

- **Call volume is handled by Live Agents or through a Natural Language Interactive Voice Response (NL IVR) system.**
  - Dependent on type of call

- **Customer Service Representatives take calls about emergencies, power outages, meter readings, billing issues, establishing or transferring electric service and various other matters.**

Prior Call Volume Forecast

- **Forecast for <u>Monthly</u> Live-Agent Call Volume for next twenty-four months**

- **Forecast for <u>Daily</u> Live-Agent Call Volume for next ninety days**

- **Forecast used for:**
  - Input into estimating the long-term staffing that we require in our call centers. These are monthly FTE requirements going out 12 months.
  - Primary input for the short-term labor capacity management. Short-term defined as weekly (includes vacations and training).
  - Input for the scheduling of agents/contractors within our call center.
  - Forecasted call volumes used as a weighting factor for the forecast / projections of our average speed of answer and service level performance factors within our call centers.

- **A manually intensive process**

Correlation of Drivers to Total Call Volume

ABANDONED
- Monthly: 0.508
- Weekly: 0.516
- Daily: 0.605

BILL
- Monthly: 0.687
- Weekly: 0.585

BILLING EXCEPTIONS TOTAL
- Monthly: 0.678
- Weekly: 0.618

DISCONNECT CALLS ATTEMPTED
- Monthly: 0.714
- Weekly: 0.649

DISCONNECT CALLS CONNECTED
- Monthly: 0.742
- Weekly: 0.668

REMINDER CALLS ATTEMPTED
- Low significance

REMINDER CALLS CONNECTED
- Low significance

IMPL
- Monthly: 0.649
- Weekly: 0.654

MIMO
- Monthly: 0.758
- Weekly: 0.696
- Daily: 0.594

MOVE-IN
- Monthly: 0.704
- Weekly: 0.607

MOVE-OUT
- Monthly: 0.768
- Weekly: 0.626

NOTF
- Monthly: 0.639
- Weekly: 0.633

NUMBER OF TERM NOTICES
- Monthly: 0.783
- Weekly: 0.681

READ-RATE
- Monthly: 0.915
- Weekly: 0.666

Note: R values were only highly significant (above 0.75) at the monthly level; there are much lower correlations between each respective data factor and total call volume for weekly/daily timeframes.

| Date Features	Date Features	Additional Call Features
Week number	Day of week	Abandonment Rate
Moratorium flag	FE Holiday	Billing exceptions
Workday flag	Day after Holiday	Billing implausibles
Weekend flag	Day before Holiday	Total bills issued
Business days from 1st of month	Day of week	Average bill amount
Business days from last of month	Month	Meter read rate
Non-Monday holiday flag	Year	Number of term notices
Moving averages	Workdays in week	Number of move ins/move outs

Historical Live Agent Call Volume

- **Though there were interesting findings within the correlations to other features, we started our model simply using historical call volume and date dimensions**

- **Gradient Boosted Trees model for Daily forecast**

- **Holt-Winters time series model for Monthly forecast**

Actual 5-Year Call Volume

Year-Over-Year Variance of Call Volume
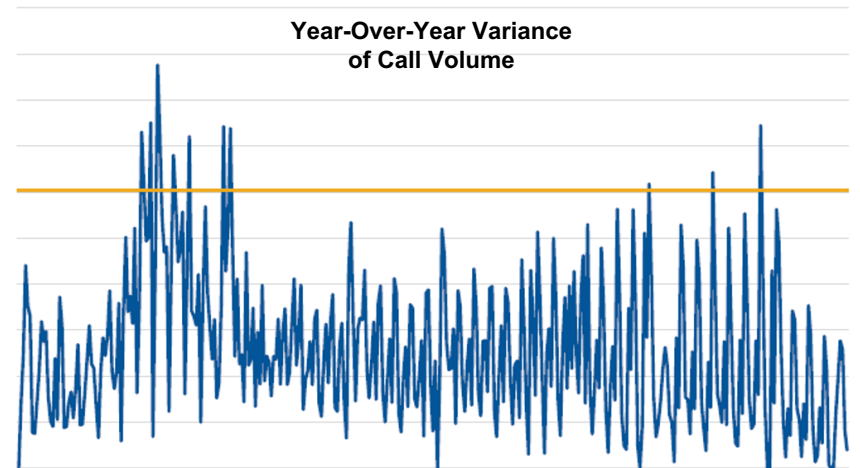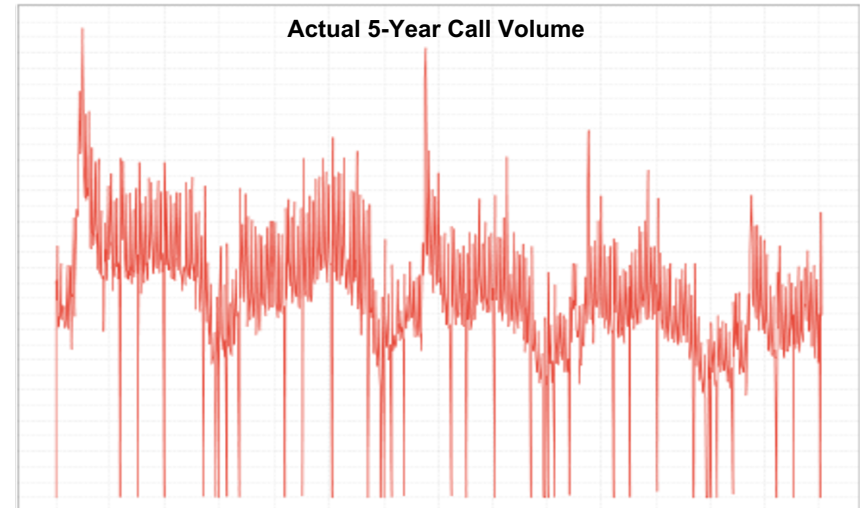
* scale not provided

FirstEnergy

Forecasting Time Series Data
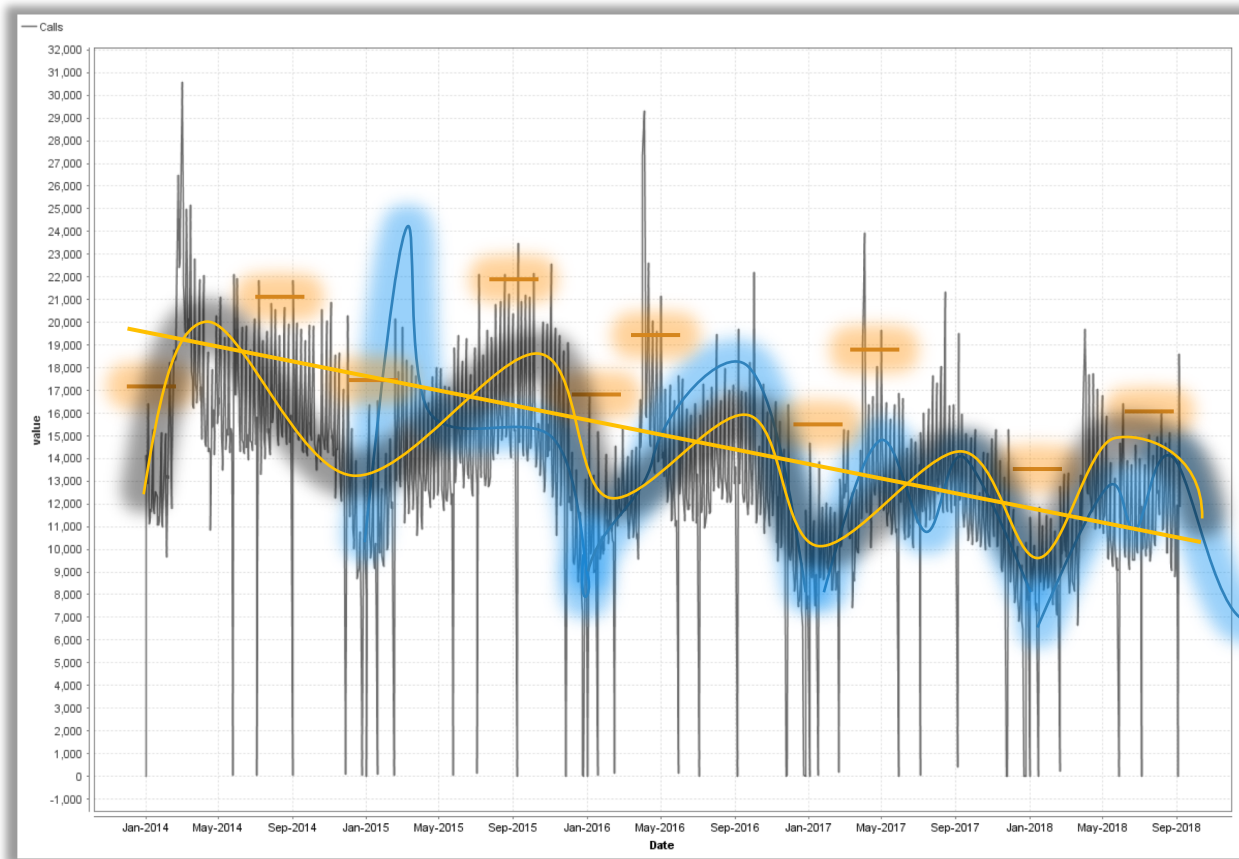
Machine Learning
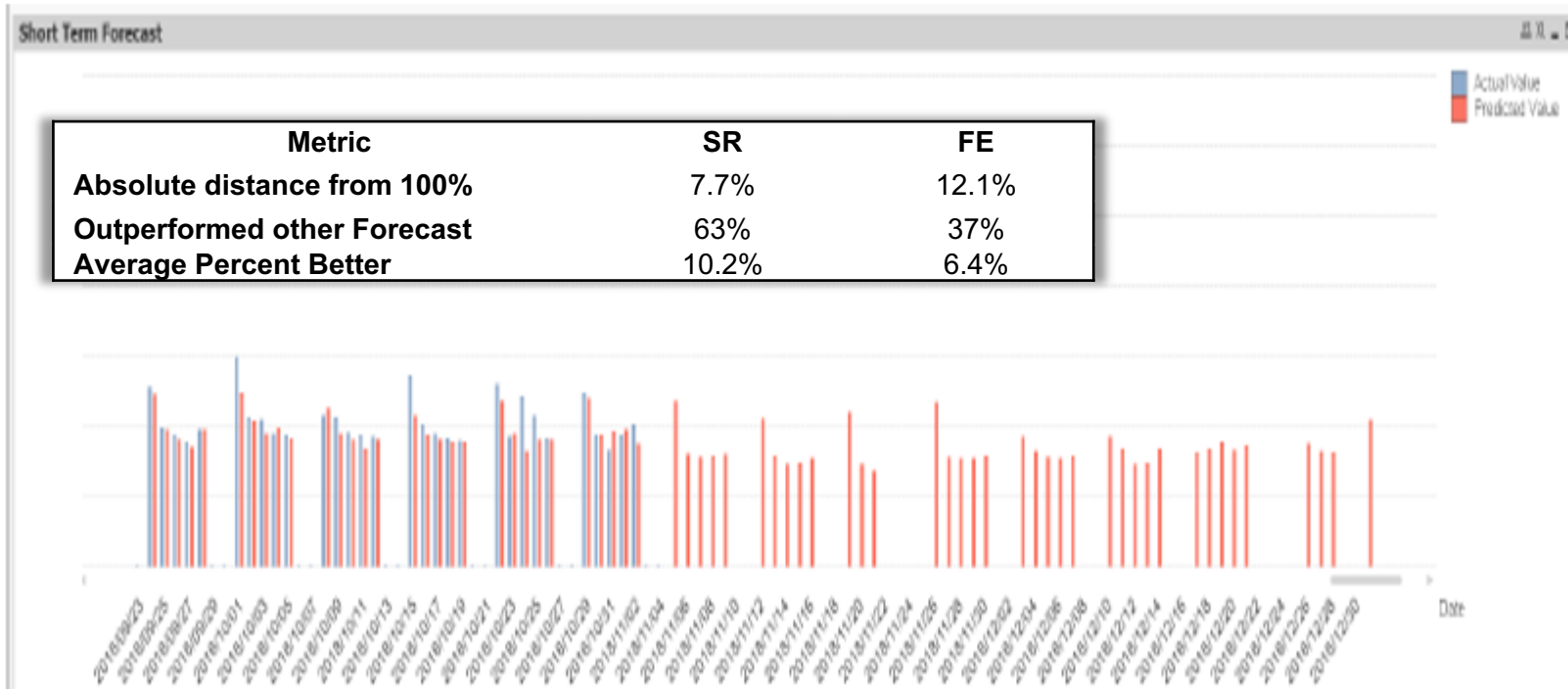
Guessing/Heuristics

% Change From Last Year

Statistical Methods

Most other forecasting methods only take into account single points of data however ML can leverage much more as inputs

| Additional Features |
| --- |
| Day of week |
| Holidays |
| Weather |
| Pricing |
| Other forecasts |
| Predictions from other predictions |

Short-term Model Results

■ **Daily call volume forecast, with blind test results shown over six weeks of 2019 and forecast results shown for following eight weeks.** (Gradient Boosted Trees)

Metric	SR	FE
Absolute distance from 100%	7.7%	12.1%
Outperformed other Forecast	63%	37%
Average Percent Better	10.2%	6.4%

Long-Term Model Results

■ **Monthly call volume forecast, with blind test results and confidence intervals going forward.** (Holt-Winters)
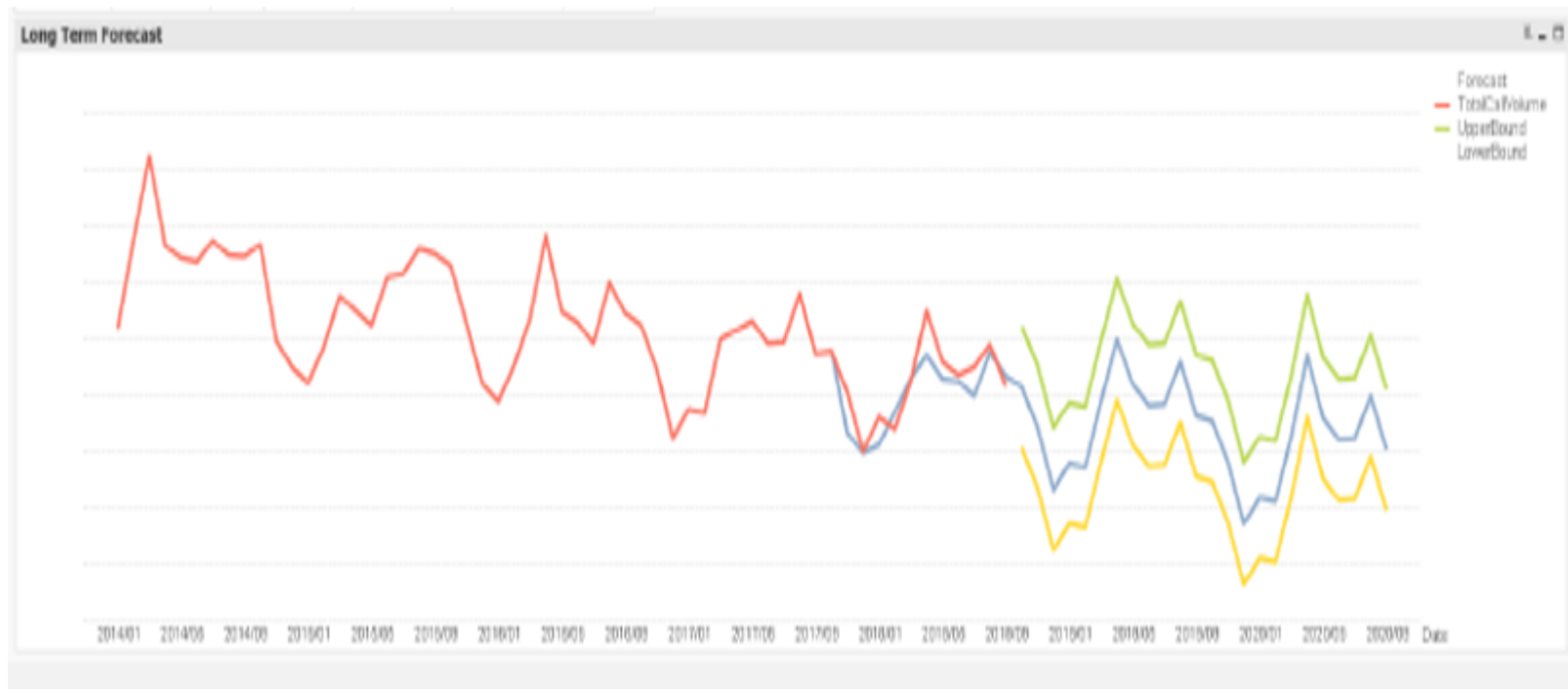
Process Improvement Achieved (~six weeks)

Previous Methods	**RapidMiner**

Previous Methods

- **Long-term model prepared annually, updated quarterly**
  - Manual effort

- **Short-term model derived from LT results quarterly**

- **Results converted to QlikView report**

RapidMiner

- **Long-term model can be run on demand**
  - Currently run quarterly

- **Short-term model is now scheduled to run weekly**

- **Model output ported to QlikView report**

- **Improved accuracy of blind test period results by 50%**

Challenges

- **Defining "Production"**
  - FirstEnergy defines a "Production Process" as a one that has a fully automated data pipeline, needs no ongoing alterations and has been previously moved from Dev to Staging, where it was reviewed.
  - Only IT has access to our Production environment.
  - This creates a challenge where no one wants to move projects into production because they lose the ability to…

- **Ownership**
  - Even with the best hand-off possible, long-term ownership of processes is challenging.
  - If the Process is owned by the Business Unit, they rarely have the technical expertise to troubleshoot issues, if it is owned by IT in Production they do not have the expertise to understand issues.

Consultant as a Catalyst

- **Since this was our first project with RapidMiner, we choose to engage Clarkston Consulting to give us a "step-up".**

- **Using a third-party partner for our first project with RapidMiner had important benefits**
  - Understanding of an End-to-End, fully automated solution
  - Asking the right question (or re-asking a question, such that it is answerable and actionable)
  - Accountability
  - Comprehensive documentation
  - Gaining additional visibility/hype into the value of Rapidminer
  - Identification of infrastructure deficiencies
  - Identification of knowledge gaps
  - Learning some technical "tricks" (like loop-apply)

Challenge Using Time Series

- **Windowed response variables in time series models create a particular challenge.**

  - Machine Learning models built on windowed data are usually built to forecast only one value forward

  - If we need to forecast 90 days forward, we can't just use apply model, we must use the Recall/Remember operators inside a loop
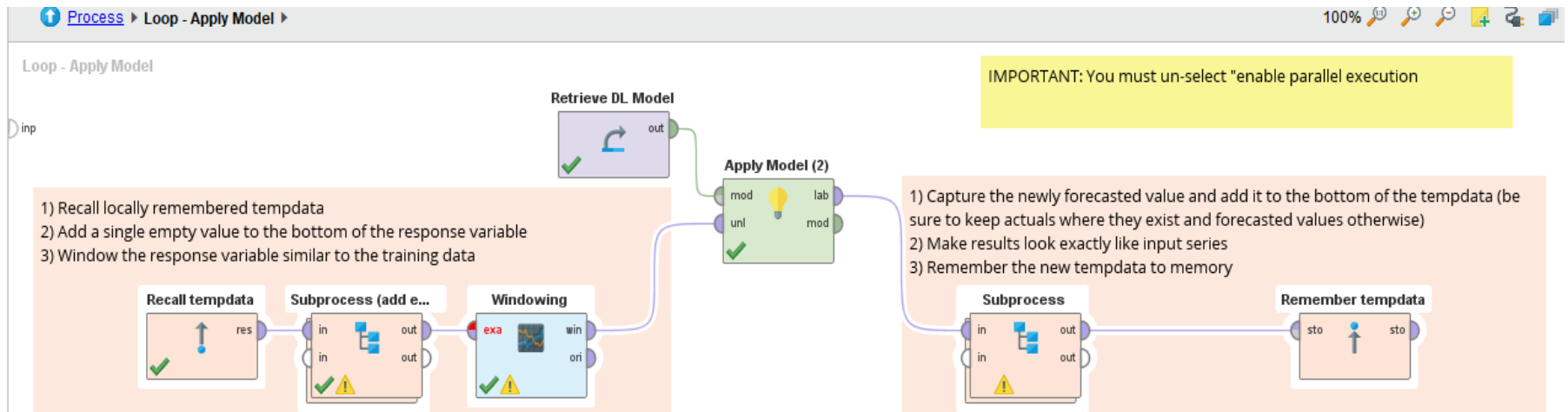
Gaps/Insights from Clarkston

- **Formalization of an FE Analytics Process**
  - Conduct Knowledge Transfer Throughout
  - Go / No-Go meeting should be part of project when changing a process
  - Don't forget about people change management

- **Consider developing Data Engineering team**
  - Data pipeline always takes longer than expected
  - Data governance and engineering needs will grow in importance as analytics grows

- **Data Space for analysts**
  - Create a "Sandbox"

- **Develop production maintenance capabilities and strategy**
  - Need a team to maintain models, data and respond to issues

- **Continue to push activating analytics**
  - Leverage RapidMiner for a different use case, like classification
  - Work with other function analytics groups using these best practices
  - Go-To resources that people can bounce ideas off of, conduct reviews, own/train processes

Next Steps

- **For Call Volume Forecast:**
  - Continue to optimize Daily Call Volume Forecast to improve accuracy
  - Produce additional forecasts for different call types
  - Include additional attributes that will be more highly correlated to individual call types
  - Rollout existing model to other call center locations

- **For FirstEnergy and RapidMiner**
  - Use this project as an example of the value of RapidMiner
  - Insert RapidMiner into other reporting and analytics processes
  - Demonstrate value to users through events, such as Hack-a-thons

- **Questions or comments?**