Predicting the Impact of Storms on Utility Customers Using Weather Analytics Models (BL14)

David J Corliss, PhD is a Senior Data Scientist at DTE Energy, a major regional electrical and gas utility based in Detroit, MI. His work in advocacy using statistics includes teaching an ASA traveling course on ethical issues in data and analytics, serving on the Committee on Scientific Freedom and Human Rights and communications chair for the Justice, Equity, Diversity, and Inclusion outreach group, and the steering committee of the Statistics section of the American Association for the Advancement of Science.

Predicting the Impact of Storms on Utility Customers Using Weather Analytics Models BL14

David J Corliss, PhD

Weather Analytics Model: Project Overview

Overview: Project Vision Statement

Deliver a product that can provide predictability on a regional approach so that DTE can resolve outage issues and apply resources where needed in a cost-effective way

Outage Predictability

Hourly prediction of events by Service Center

Identify Areas of Greatest Impact

Regional damage multiplier prediction

Resource Management

Predict number of resources needed

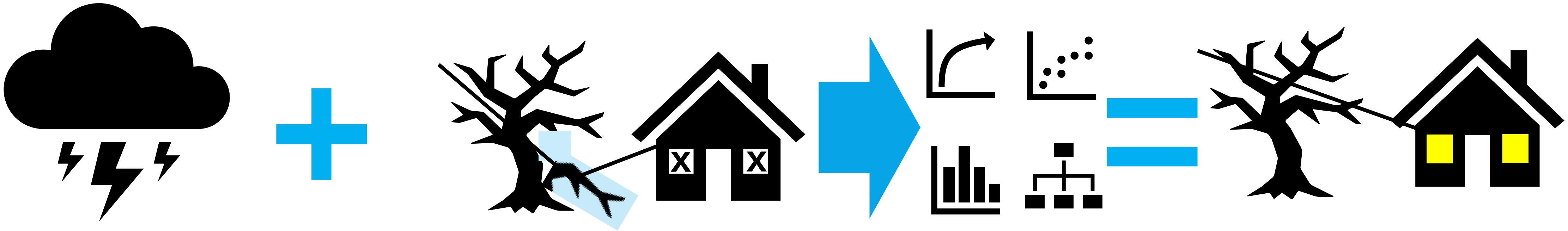
Overview: Concept Flow Chart

Leverage Publicly Available High Quality Weather Forecast Data from NOAA

Match to DTE Outage History

Apply Multiple Modeling Methods

Local Outage Predictions by Day



Overview: Business Model

Everything depends on a

close partnership between

three key stakeholders

Meteorology Team

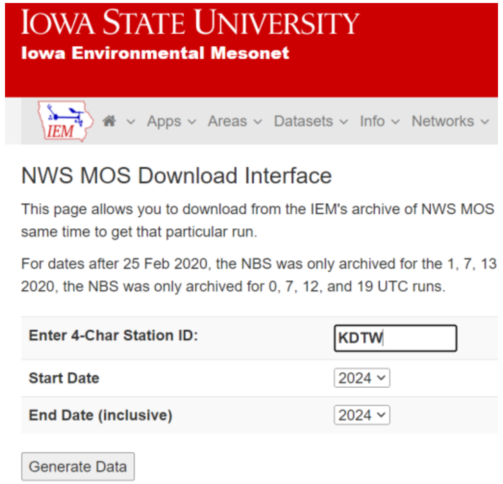
Data Science Team

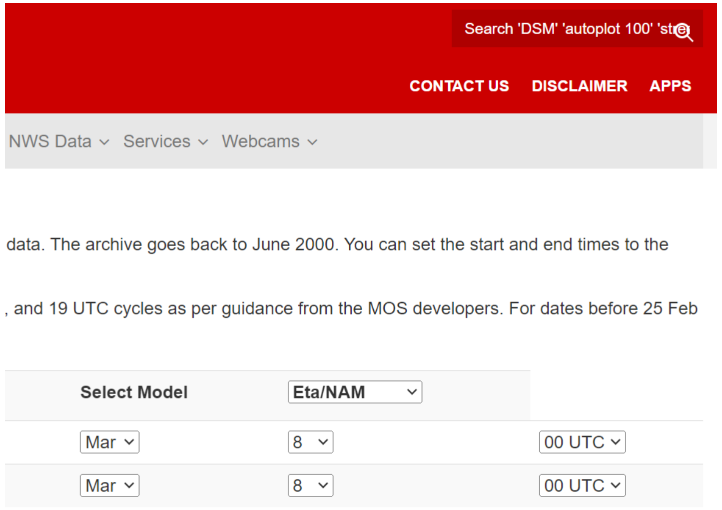
Accurate, Reliable, Timely, Local Outage Predictions Drive Business Value

Business Team

Weather Analytics Model:
Data Sources and Feature Importance

Data: Iowa State Weather Forecast Archive





Available Data	
Max/Min Temp [F]	Freezing Rain Probability [%]
2m Air Temperature [F]	Snowfall Probability [%]
2m Dew Point [F]	Precipitation Type
Cloud Coverage	Sky Coverage [%]
10m Wind Direction [deg]	Wind Gust [kts]
10m Wind Speed [kts]	3 Hr Thunderstorm Prob [%]
6 Hr Probability of Precipitation [%]	Probability Freezing Rain [%]
12 Hr Probability of Precipitation [%]	Probability of Snow [%]
6 Hr Quantitative Precip Forecast	Probability of PL [%]
12 Hr Quantitative Precip Forecast	Probability of Rain [%]
6 Hr Thunderstorm Probability [%]	0-6 km Vertical Wind Shear
12 Hr Thunderstorm Probability [%]	Snow Level
Categorical Snow	I06
Categorical Ceiling Height	Lowest Cloud Base
Categorical Visibility	Significant Wave Height
Obstruction to Vision	

Physics-Informed Artificial Intelligence

The integration of the principles, equations, and constraints of physical science to inform decision science, increase accuracy, and optimize processes and results

Examples

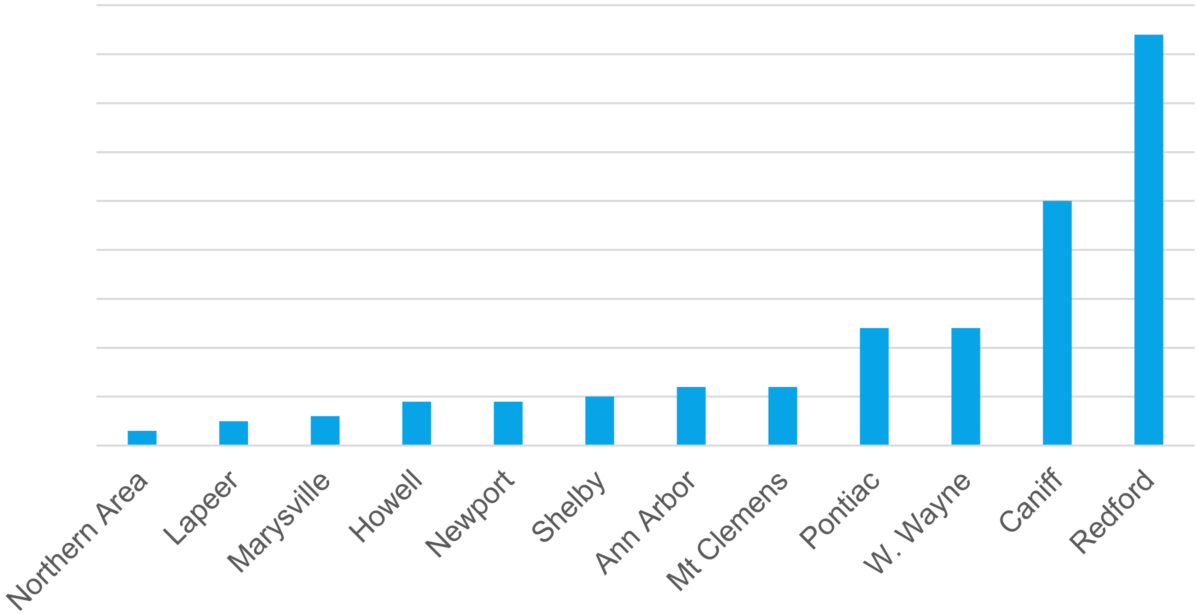
- A Application of atmospheric physics in weather forecasting algorithms
- B Analysis of the equations of motion as system constraints
- C Inclusion of energy transfer mechanisms in AI describing physical processes
- D Estimating system and algorithm performance characteristics based on physical limitations

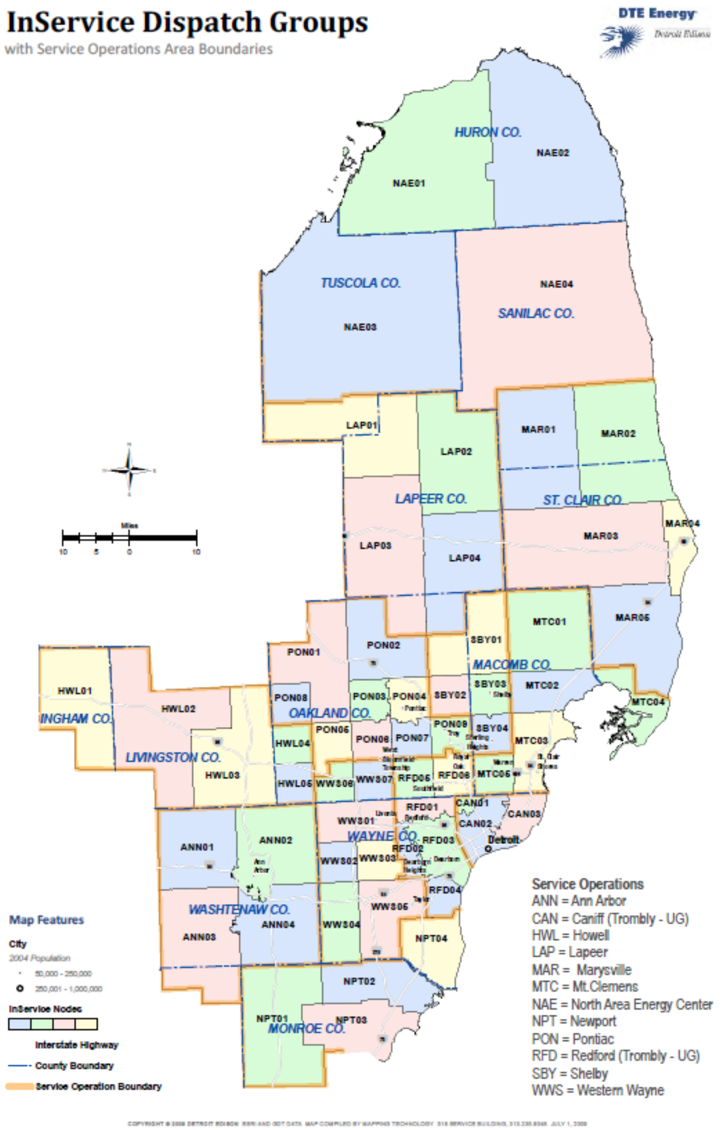
Model Predictions: Regional Differences

Top Factors Driving Regional Differences

- Number of customers in each area
- Local differences in the severity of weather events
- Close to the lakes vs. inland
- Seasonal changes have different timing
- Land use - buildings vs houses vs farms

Most Common Number of Outage Events Per Day





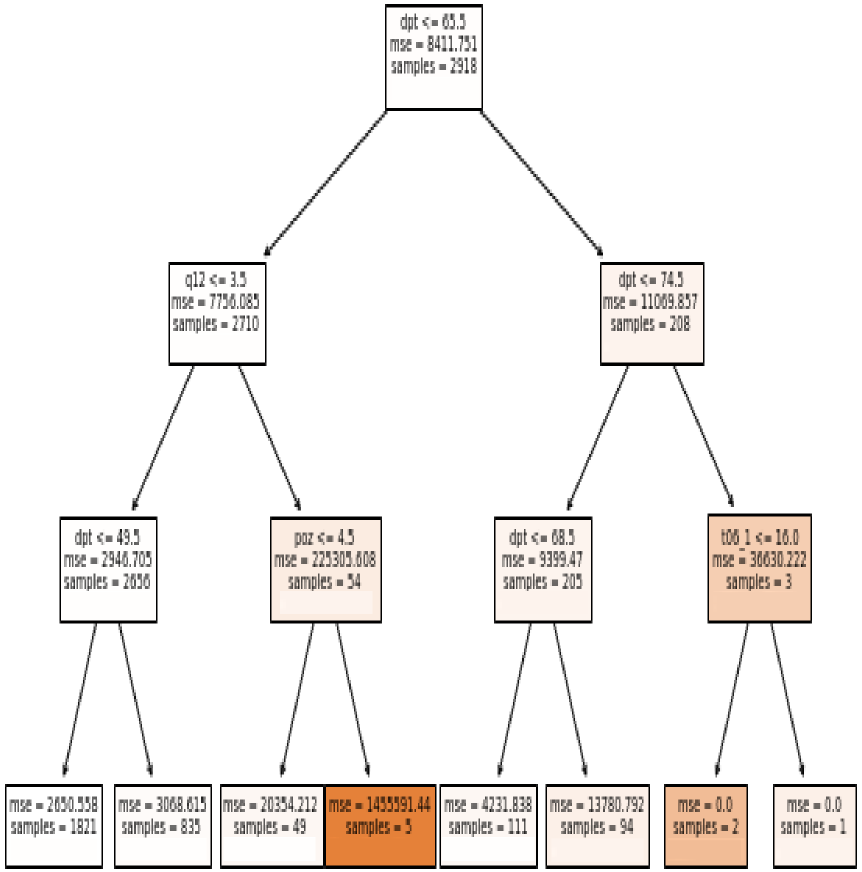
#MWSUG2024 #BL65

Model Predictions: Informative Features

Factors included in one or more local models

- **Geographic Location**
- **Seasonality**
- **Temperature**
- **Dew Point**
- **Difference between Temperature and Dew Point**
- **Precipitation: % chance, quantity next 12 hours**
- **% Chance of Thunderstorm Next 6 or 12 Hours**
- **Probability of Wind Gust Over 25 / 35 / 45 mph**
- **Wind Direction**
- **Visibility**
- **Intensification: an increase in severity of weather between the 72-hour and 48-hour forecasts**

Decision Tree identifying important forecasting fields



MWSUG 2024
WAUKESHA, WI
LEAD THE WAY AND PLUG INTO SAS!

Analytic Methods

Model Predictions: Informative Features

Suite of Models: Summary of Algorithms

Regression Methods: Predicts the Expected Number of Outages in a Day

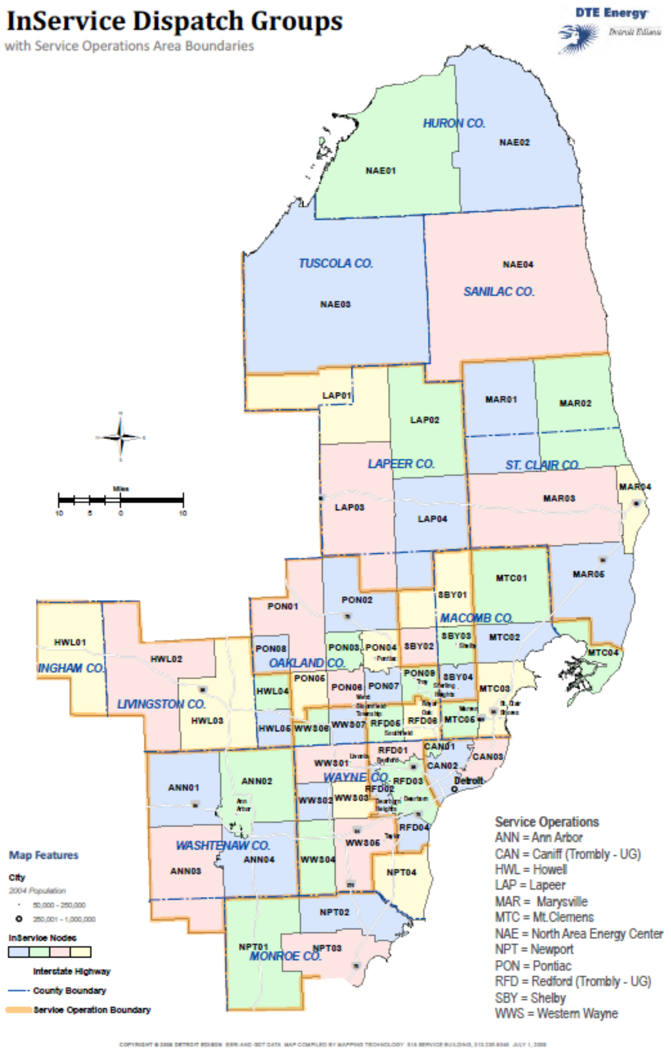
- Multiple Linear Regression *
- RandomForest *
- Neural Networks - CNN and LSTM *
- XGBoost

Classifier Methods: Predicts Weather Impact by Category – Normal, High, Storm, etc.

- Logistic Regression *
- Decision Tree
- KNN
- SVM

*** Rashomon Set**

Geo-Spatial Modeling: Localization and Boosting

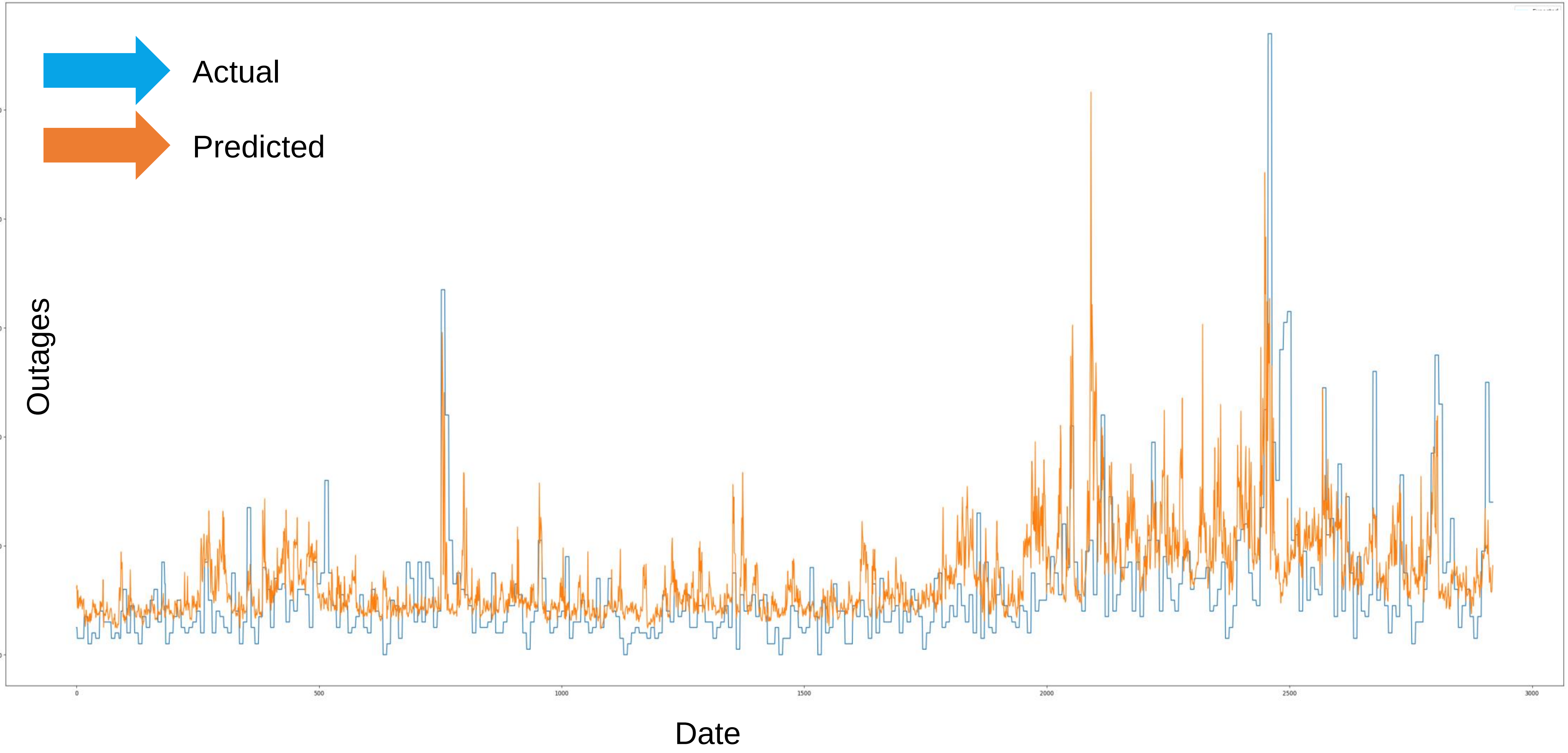


A. Due to regional differences, each DTE Service Center gets its own model

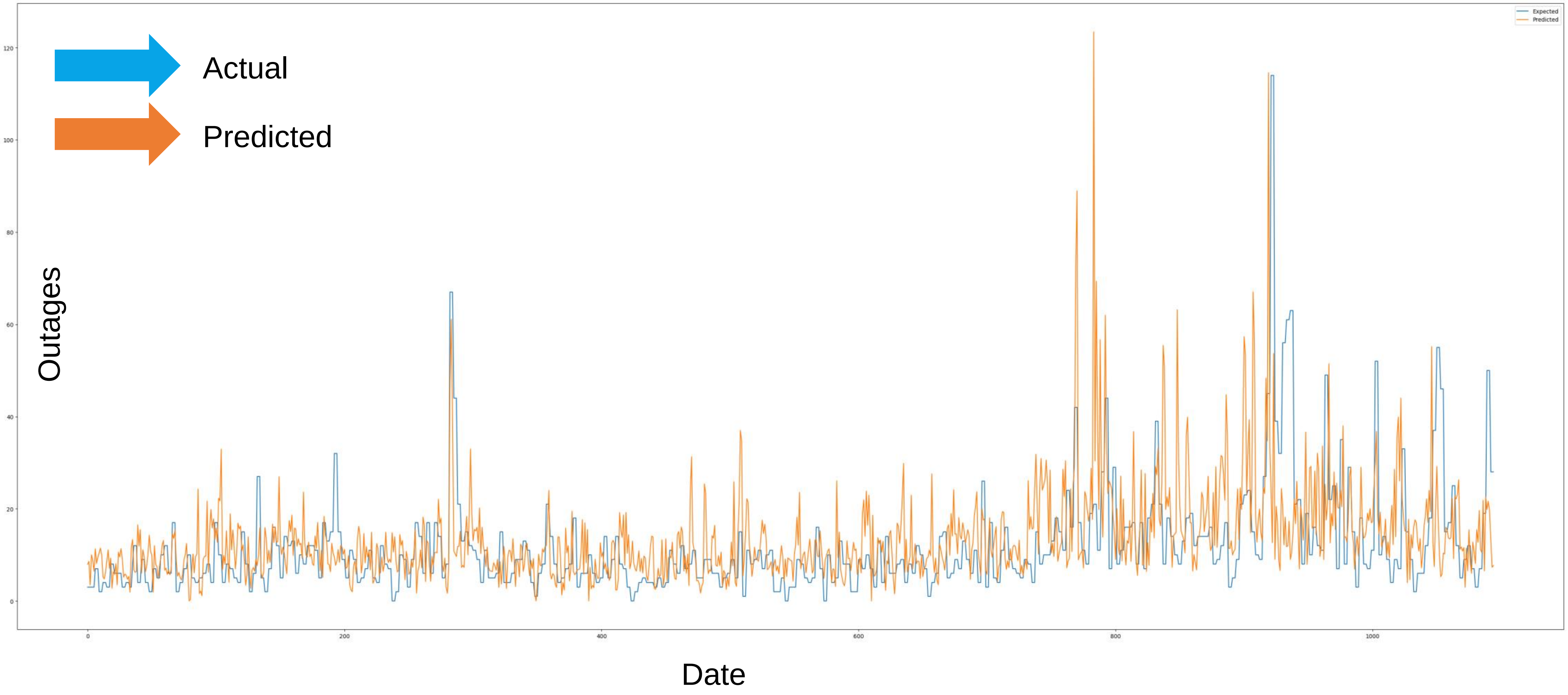
B. Boosting: numbers for each area are added for an overall total, reducing the impact of noise

C. Localization and boosting results in more accurate predictions

Daily Predictions: RandomForest



Daily Predictions: Neural Network

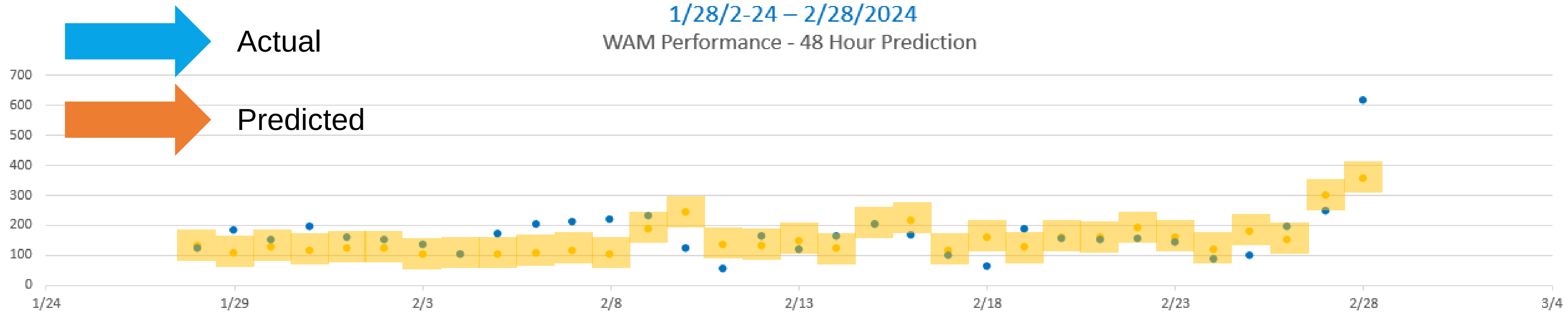


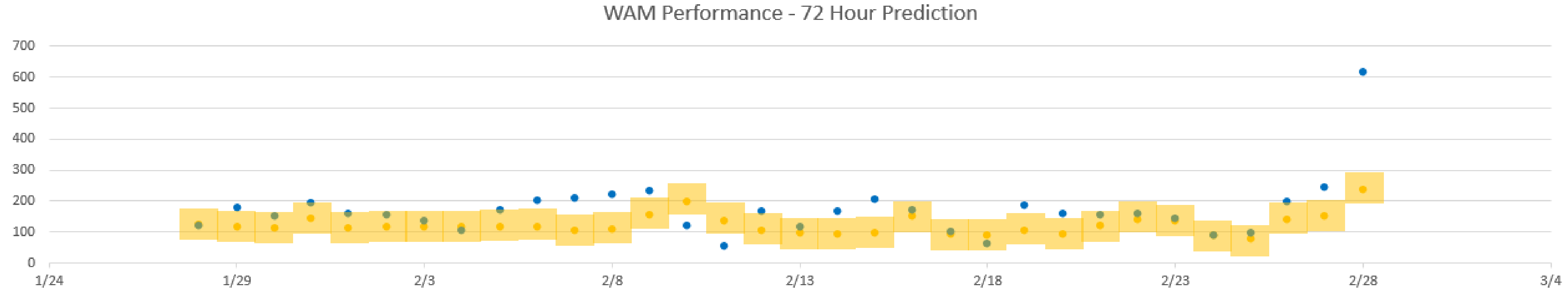
#MWSUG2024 #BL65

Predictive Performance and Monitoring

Model Performance Monitoring and Comparison

1/28/2-24 – 2/28/2024



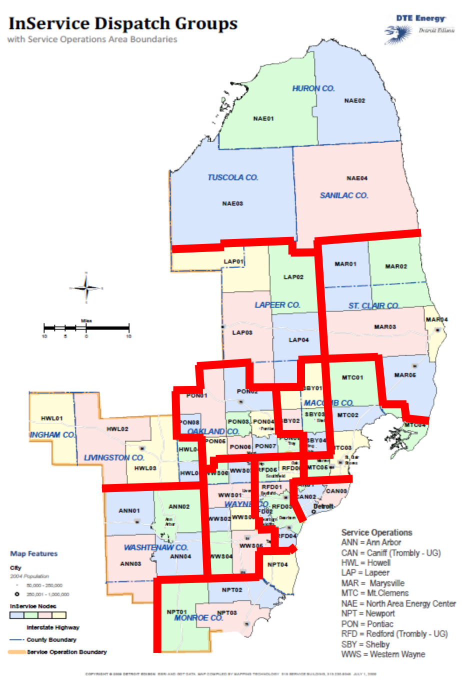


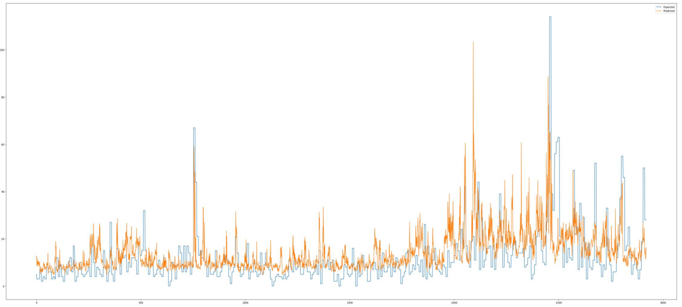
Model Accuracy

Class (Normal, High Normal etc.): 72-hour: 93.8% 48-Hour: 100%
Count (Within 50 of Actual): 72-hour: 53.1% 48-Hour: 75.0%

Conclusions

Best Practices for Expert Witnesses





1. Predicts number of outages 48 and 72 hours in advance using weather forecast data and historical outage counts
2. Multiple model techniques, including regression, classification, and machine learning methods
3. A separate version of each model is developed for each local area and then added together
4. CI/CD in analytics: models are constantly monitored and improved

#MWSUG2024 #BL65

Questions

?

Thank You!

- Contact Information
- David J Corliss, PhD
- Grafham Analytics
- E-mail: davidjcorliss@gmail.com
- LinkedIn: linkedin.com/in/davidjcorliss/

Trademark Citation

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.