TAP TO GO BACK TO KIOSK MENU

SAS® GLOBAL FORUM 2020

MARCH 29 – APRIL 1
WASHINGTON, DC

USERS PROGRAM

PREDICT FREEZES IN NATRUAL GAS PIPELINES

TAP TO GO BACK TO KIOSK MENU

UNIVERSITY OF Central Oklahoma

Bingi Arnold Kanagwa M.S, Ho -Chang Chae, PHD

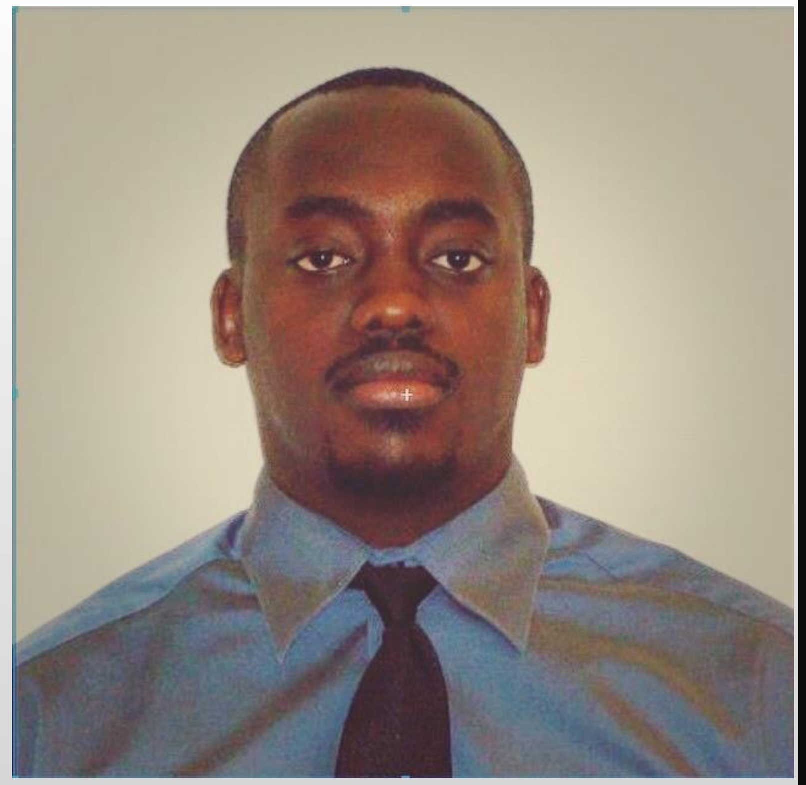
University of Central Oklahoma

Abstract
Introduction
Methods
Results 1
Results 2
Conclusion

In 2018, natural gas was the largest source of energy production in the United States, accounting for 31.8% of all energy production. Midstream companies are responsible for transportation and processing of natural gas. One of the biggest problems facing midstream companies is meter freezing. A meter freeze is caused by the presence of natural gas hydrates (a high-pressure form of frozen water and natural gas molecules) in the tap valves, gauge lines, or manifold so that the differential sensing element is effectively no longer connected to the flow line. Because one side of the manifold is likely to freeze before the other, small changes in line pressure on either side of the orifice can create large changes in the indicated differential pressure record. The differential may read either high or low as a result of meter freeze, and the effect may cause substantial measurement error.

Currently, data analysts spend hundreds of man-hours looking through data for possible meter freezes, and this manual process is time-consuming, tedious, and error-prone. Thus, the goal of our project is to use advanced machine learning models in SAS and Python to identify meter freezes. In SAS Enterprise Miner we used Logistic Regression, Neural Network, Decision Tree, Gradient Boosting, and the HP Forest nodes. With Python scikit-learn library we ran Logistic Regression, Random Forest, Decision Tree, Ensemble, and XG Boost Models. SAS Enterprise Miner Gradient Boosted Model outperformed the rest based on the overall F1 score.

The best model is able to correctly classify meter freezes, and it is saving a midstream company millions of dollars in revenue and man-hours.



Bingi Arnold Kanagwa

PREDICT FREEZES IN NATURAL GAS PIPLELINES

Bingi Arnold Kanagwa M.S, Ho Chang Chae, PHD

University of Central Oklahoma

Natural Gas Pipelines and Meter Freezes

- Natural gas in the United States was the nations largest source of energy production in 2018
- This represents 31.8% of all energy production in the country
- In 2018, the United States produced 32.7 trillion cubic feet of marketed natural gas
- Average well head values is $2.66 per thousand cubic feet
- Total well head value of $85.3 billion
- Meter freezes are one of the biggest sources of lost revenue for natural gas companies.

  - Meter freezes are caused by the presence of natural gas hydrates (a high pressure form of frozen water and natural gas molecules) in the tap valves, gage lines, or manifold.
  - Then, the differential sensing element is effectively no longer connected to the flow line.
  - Meter freezes cause substantial measurement error which leads to tens of millions of dollars in lost revenue for midstream companies

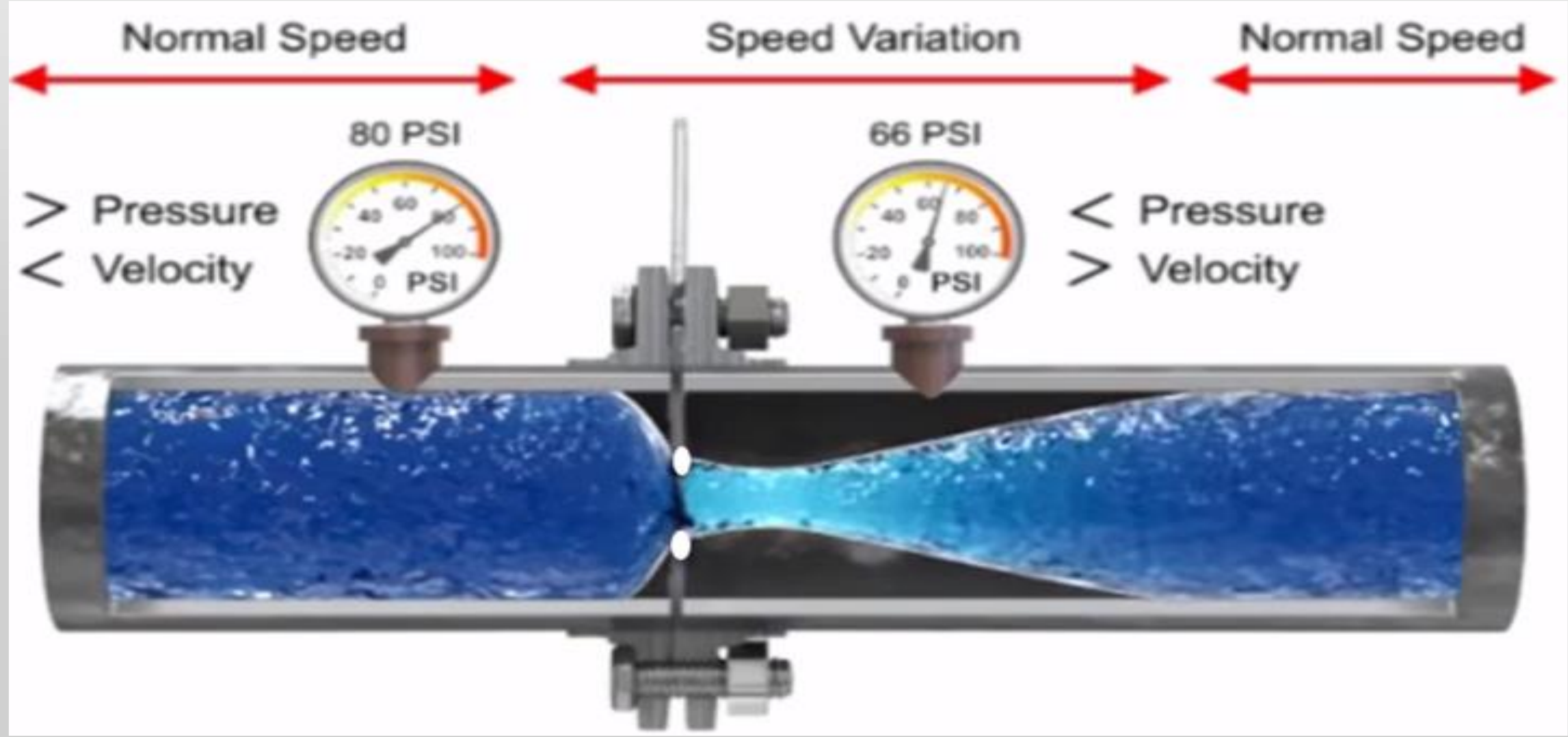


Illustration of a meter freeze

Objective

- Build machine learning models to quickly identify freezes
- Improve accuracy by reducing human error
- Save time and money

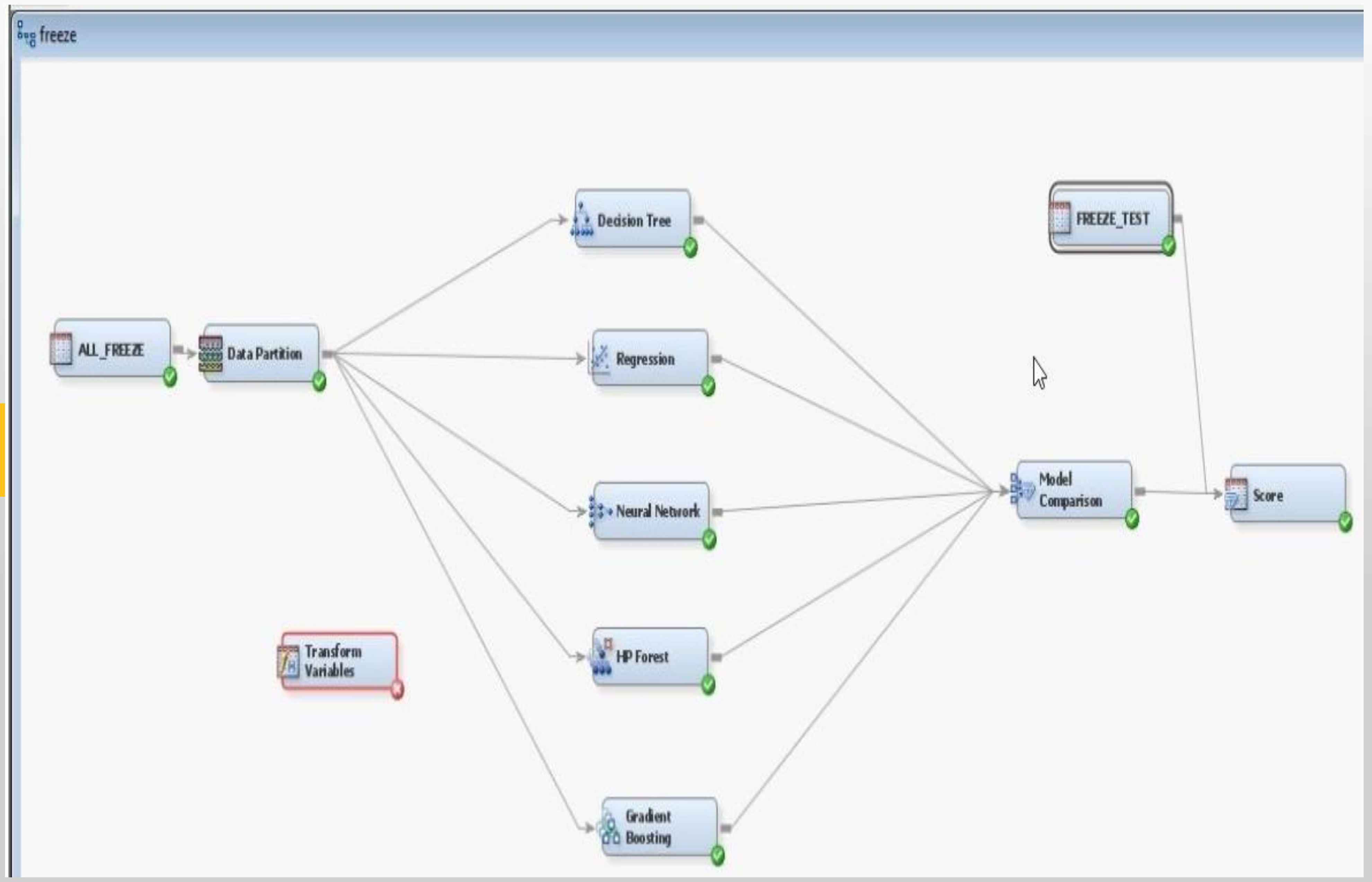
PREDICT FREEZES IN NATURAL GAS PIPLELINES

University of Central Oklahoma

Bingi Arnold Kanagwa M.S, Ho Chang Chae, PHD

University of Central Oklahoma

- Abstract
- Introduction
- **Methods**
- Results 1
- Results 2
- Conclusion



SAS Enterprise Miner

- Decision Tree
- Logistic Regression
- Neural Network
- HP forest
- Gradient boosted (selected model)

Scikit-learn

- Logistic Regression
- Random Forrest
- Decision Tree
- Extreme Gradient Boosted (selected model)
- Ensemble Model

```
#model with no transformations
def my_model_nt (data ):
    data_2=data.set_index('Date')
    data_3=data_2.dropna()
    data_4=data_3.drop(['Number','Name'],axis=1)
    predictions=xgb_clf.predict(data_4)
    probabilities=xgb_clf.predict_proba(data_4)
    score=probabilities[:,1]
    data_3['predictions']=predictions
    data_3['score']=score
    data_6=data_3[data_3['predictions']==1]
    data_7=data_6.drop(['Differential','Pressure','Temperature','Flow Duratic
    data_7.to_csv('arnold_freeze_xgb.csv')
    return(data_7)
```

PREDICT FREEZES IN NATURAL GAS PIPLELINES

University of Central Oklahoma

Bingi Arnold Kanagwa M.S, Ho Chang Chae, PHD

University of Central Oklahoma

	SAS Enterprise miner		
Model Description	precision	recall	F1 score
Decision Tree	0.8750	0.8077	0.84
Regression	0.6875	0.4231	0.523809524
Neural Network	0.8214	0.8846	0.851851852
HP Forest	0.8800	0.8462	0.862745098
Gradient boosting	0.8846	0.8846	0.884615385

SAS Enterprise Miner

- Precision (accuracy of positive results)
- Recall (ratio of positive instances correctly detected by the classifier)
- F1 score is a combination of the two measures
- Gradient Boosted model performed the best

Scikit-learn

- Random Forest model had the best precision
- Gradient Boosted model had the best over-all score based on F1 score.

	Sci-kit learn		
Model Description	precision	recall	F1 score
Decision Tree	0.6296	0.7083	0.666635294
Regression	0.7142	0.625	0.666629331
Random Forest	0.8947	0.7083	0.79066252
Gradient boosting	0.8333	0.8333	0.8333
Ensemble	0.8888	0.6666	0.761828571

SAS Enterprise Miner vs. Scikit-learn

1. Overall, models created with SAS Enterprise Miner performed better than models developed using Scikit-learn

PREDICT FREEZES IN NATURAL GAS PIPLELINES

UNIVERSITY OF Central Oklahoma

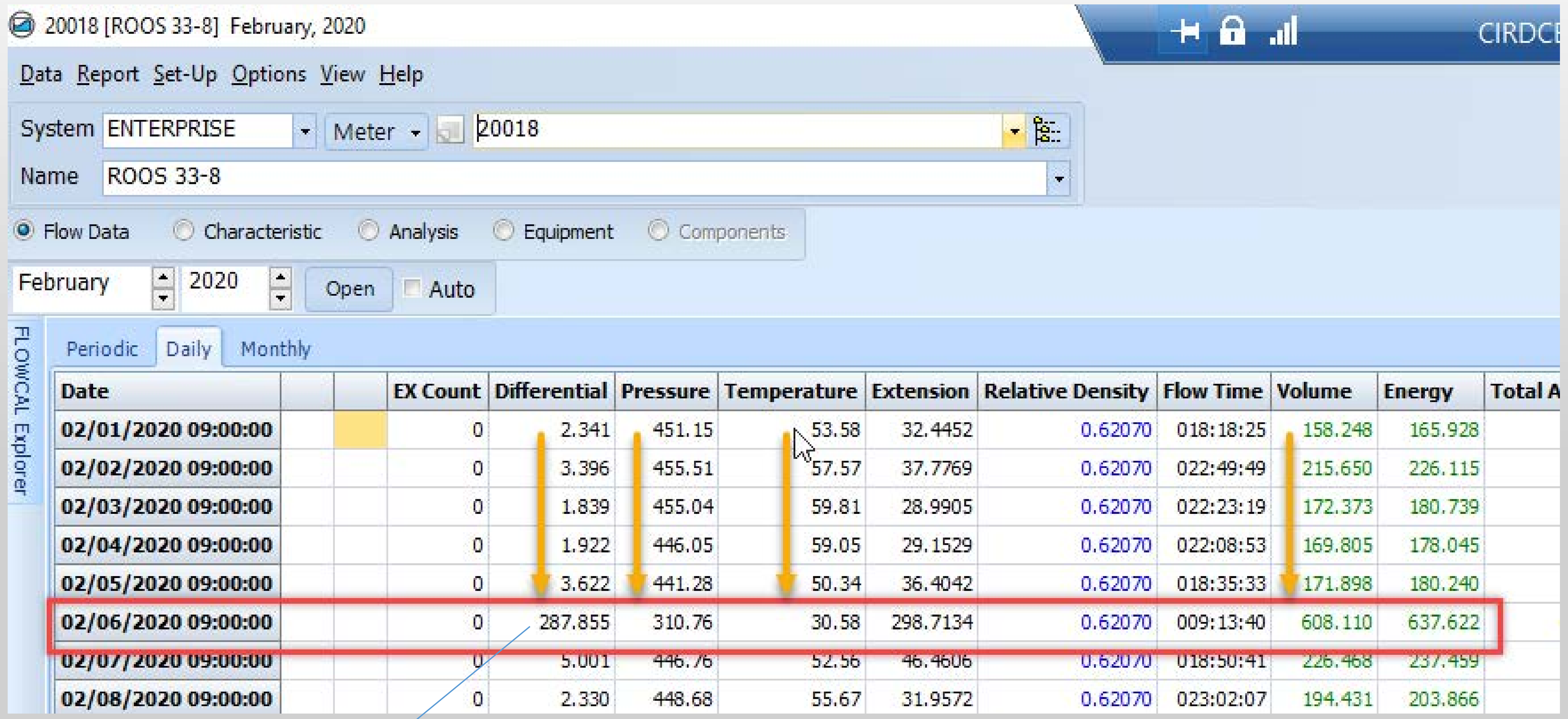
Bingi Arnold Kanagwa M.S, Ho Chang Chae, PHD

University of Central Oklahoma

Scoring and Implementing Selected Model

- Abstract
- Introduction
- Methods
- Results 1
- **Results 2**
- Conclusion

- Results from gradient boosted model was used to identify freezes
- Machine learning significantly reduced the amount of time taken to classify freezes
- Less than 3 seconds to classify over 13000 meters
- Human analysts would require at least three work days.



Date			EX Count	Differential	Pressure	Temperature	Extension	Relative Density	Flow Time	Volume	Energy	Total A
02/01/2020 09:00:00			0	2.341	451.15	53.58	32.4452	0.62070	018:18:25	158.248	165.928	
02/02/2020 09:00:00			0	3.396	455.51	57.57	37.7769	0.62070	022:49:49	215.650	226.115	
02/03/2020 09:00:00			0	1.839	455.04	59.81	28.9905	0.62070	022:23:19	172.373	180.739	
02/04/2020 09:00:00			0	1.922	446.05	59.05	29.1529	0.62070	022:08:53	169.805	178.045	
02/05/2020 09:00:00			0	3.622	441.28	50.34	36.4042	0.62070	018:35:33	171.898	180.240	
02/06/2020 09:00:00			0	287.855	310.76	30.58	298.7134	0.62070	009:13:40	608.110	637.622	
02/07/2020 09:00:00			0	5.001	446.76	52.56	46.4606	0.62070	018:50:41	226.468	237.459	
02/08/2020 09:00:00			0	2.330	448.68	55.67	31.9572	0.62070	023:02:07	194.431	203.866	

Date	Number	Name	score
2/3/2020 9:00	3931	HUDGINS #17-2	0.5502276
2/6/2020 9:00	3931	HUDGINS #17-2	0.6271132
2/4/2020 9:00	3951	SCHOU #1-21	0.68662345
2/5/2020 9:00	105340	BOLES FIELD CHECK	0.63042957
2/5/2020 9:00	110464	HUTCHINSON 27 H #1	0.66404
2/6/2020 9:00	110464	HUTCHINSON 27 H #1	0.91817516
2/13/2020 9:00	110464	HUTCHINSON 27 H #1	0.5946889
2/6/2020 9:00	**20018**	**ROOS 33-8**	**0.9836052**
2/6/2020 9:00	20411	HAMMETT 32-1 ALT	0.97913504
2/12/2020 9:00	208373	B S HARRISON	0.7180905
2/13/2020 9:00	208373	B S HARRISON	0.545215
2/6/2020 9:00	208382	WEATHERFORD CP	0.8604461
2/3/2020 9:00	208449	ETHANE RESIDUE	0.68537647
2/12/2020 9:00	208449	ETHANE RESIDUE	0.66393965

PREDICT FREEZES IN NATURAL GAS PIPLELINES

University of Central Oklahoma

Bingi Arnold Kanagwa M.S, Ho Chang Chae, PHD

University of Central Oklahoma

Conclusion

- SAS Enterprise Miner models outperformed Scikit-learn
- Gradient boosted models performed best in both programs
- Successful implementation of best model significantly reduced the amount of time taken to classify freeze occurrences
- Reduced number of missed freezes due to human error

MARCH 29 – APRIL 1 | WASHINGTON, DC | #SASGF