Analysis and Prediction of The Interactions between Ground Delay Programs and Ground Stops

Marc-Henri Bleu-Laine

Academic Advisor

Prof. Dimitri N. Mavris

Technical Advisor

Dr. Tejas Puranik

Georgia Tech Aerospace Systems Design Laboratory

Outline

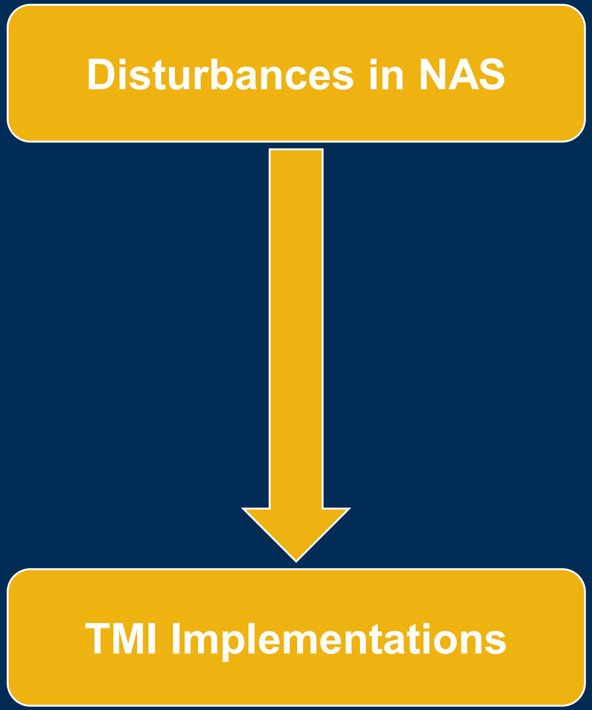
- Motivation
- Background
- Methodology
- Results
- Analysis
- Conclusion

Outline

Motivation
Background
Methodology
Results
Analysis
Conclusion

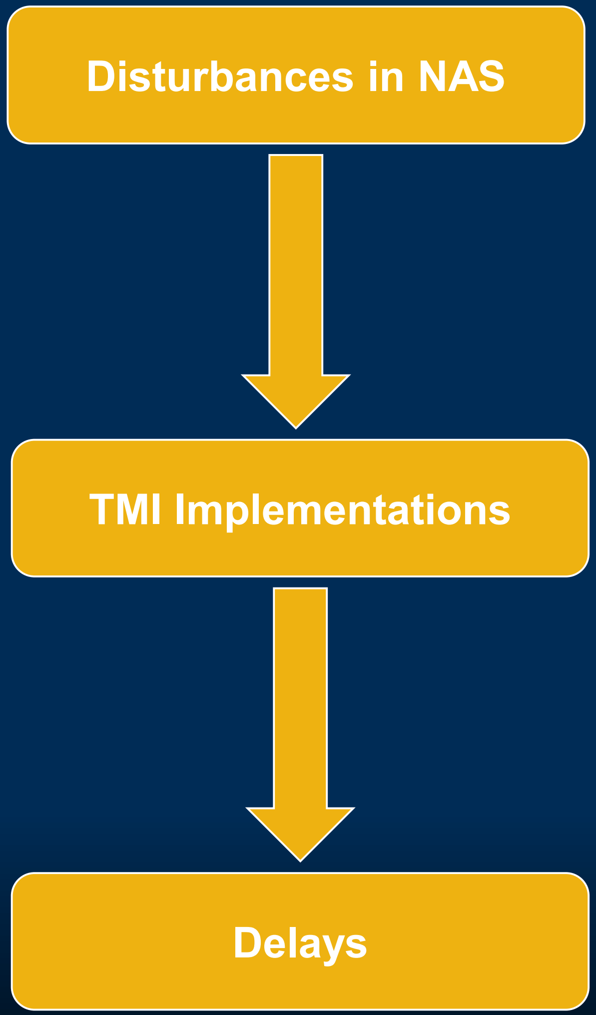
Disturbances in The NAS

- National Airspace System (NAS)
  - Network of air navigation facilities, services, airports, regulations, procedure, human resources, and material
- The NAS can be disturbed due to different reasons
  - Inclement weather
  - Runway related incidents
  - Volume constraints
  - Etc.
- Traffic Management Personnel monitors the NAS[5]
  - They observe demand and capacity at airports and can decide to implement Traffic Management Initiatives (TMI)



Traffic Management Initiatives

- TMIs are programs and tools traffic management personnel use to manage traffic [17]
  - Terminal TMIs: Regulate excess demand or lower acceptance rate at an airport
  - Enroute TMIs: Manage traffic issues in enroute environment
- Unfortunately TMI implementations create delays [11]
  - Imposes stress on air traffic controller, passengers, and the economy
  - Depending on traffic controller, more or less delays could occur [14]



Traffic Management Initiatives

- Motivation
- Background
- Methodology
- Results
- Analysis
- Conclusion

- The number of delays and their impacts can be reduced [3][4][12][13][14]
  - Previous work has focused on mainly improving the implementation of Ground Delay Programs (GDP) and Ground Stops (GS), which are terminal TMIs
- TMI interactions are common and impact the NAS differently[4]
  - From 2007-2009 35% of the days had GSs and GDP implemented on the same day[4]
  - They impact the prediction capabilities of current TMI tools
- Limited research on coincidences and interactions between GDP and GS
  - Their predictions has not been studied
  - Determining their causes can help traffic management personnel

Research Objectives

Objective 1.1 — Predict the coincidence of weather-related GS and GDP

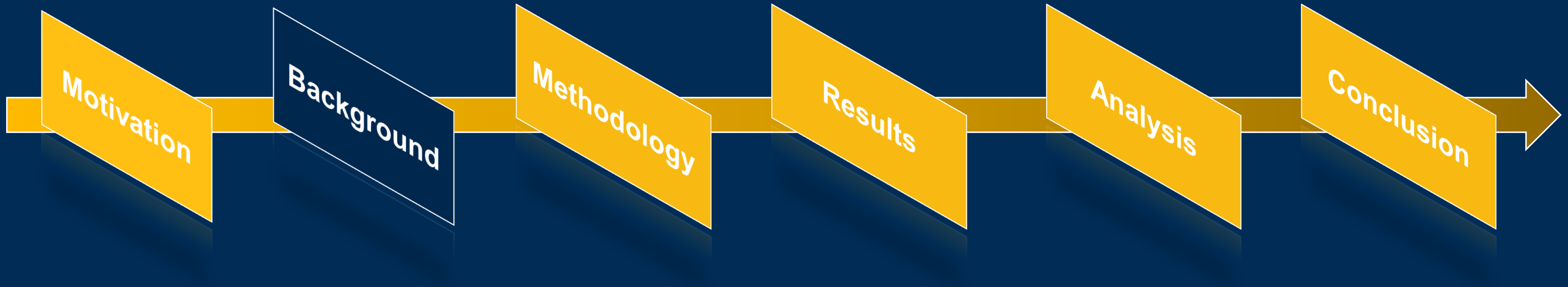
Objective 1.2 — Predict the precedence of weather-related GDP before GS, and vise versa, during their coincidence

- Reduce flight delays
- Improve the controller responses
- Improve prediction capability of TMI tools

Objective 2 — Analyze prediction models

- Discover key predictors
- Understand how can changes in predictors affects coincidences
- Understand reasons behind model decisions
- Streamline controller responses

Outline


Motivation
Background
Methodology
Results
Analysis
Conclusion

Ground Delay Programs & Ground Stops

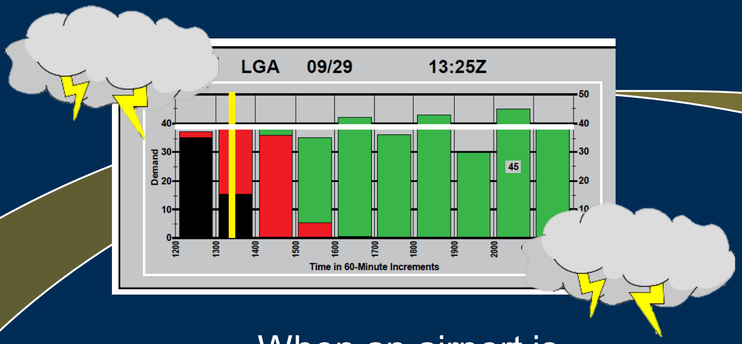
Ground Delay Programs (GDP)

- Procedure in which aircraft are delayed at their departure airport in order to manage demand and capacity at their arrival airport[5]
- They are assigned an Expected Departure Clearance Time (EDCT)

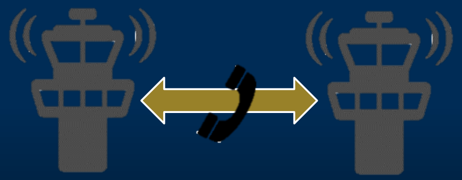
Ground Stops (GS)

- Initiative that requires aircraft that meets specific[5] criteria to remain grounded at their departing airport due to constraint at their destination
- Usually used for short-term (<2hrs)
- Durations and probability of extensions are issued
- Most restrictive TMI

GDP Scenario[5]



- When an airport is constrained, a conference call is held among impacted air traffic facilities to discuss alternatives



- Traffic Management Personnel monitors the NAS

- Advisory can be issue when GDP are proposed but not yet implemented
- Expect Departure Clearance Time (EDCT) is sent to airlines and NAS computers when the decision to implement is made
- An advisory stating that GDP is implemented is issued

GS Scenario[6]

- Significant constraint at an airport

- Traffic Management Personnel implements a GS
- A Probability of Extension (POE) and a duration time are provided when the GS is implemented
- GS are usually used reactively

- Traffic managers keep observing the situation until the end of the duration time

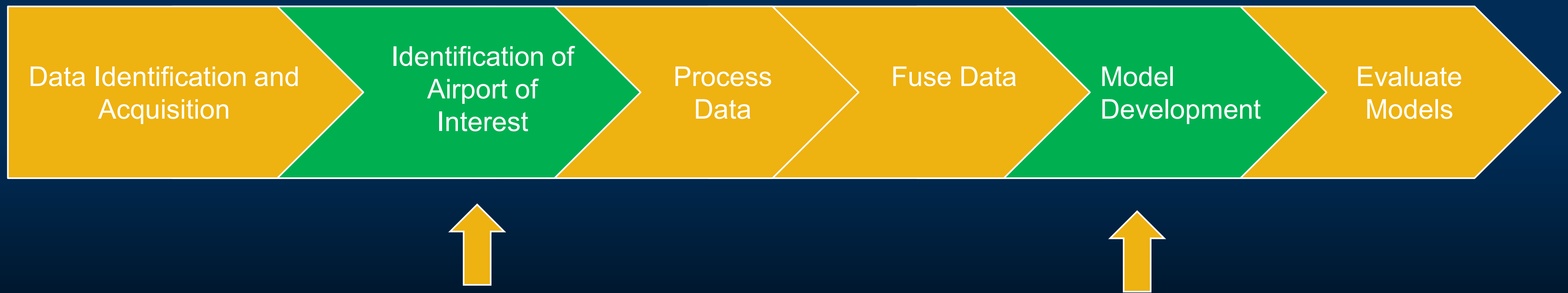
- A decision is made to pass from a GS to GDP, from a GS to another GS, or to cancel

Outline


Motivation
Background
Methodology
Results
Analysis
Conclusion

Overview of the proposed methodology[3]

- Methodology inherited from Mangortey et al[3]
  - Data Acquisition: TFMS Flight and Flow, and ASOS
  - Process Data: Conversion of data in CSV
  - Fuse Data: Fuse using date and time
  - Evaluate models: Use common metrics (Balanced Accuracy, Accuracy, Kappa Statistics, Specificity, Sensitivity)

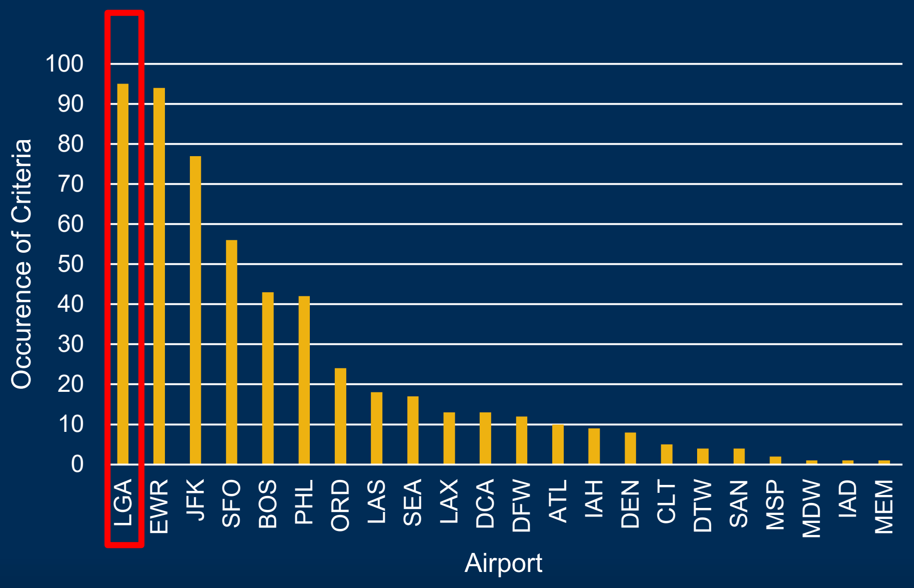


Significant Improvements and changes to inherited methodology

Identification of Airport of Interest

- 4 criteria were used to identity the airport from the TFMS Flow data
  - Number of GDPs with Preceding GS
  - Number of GDPs with Internal GS
  - Number of GS with Following GDP
  - Number of GS Internal to GDP

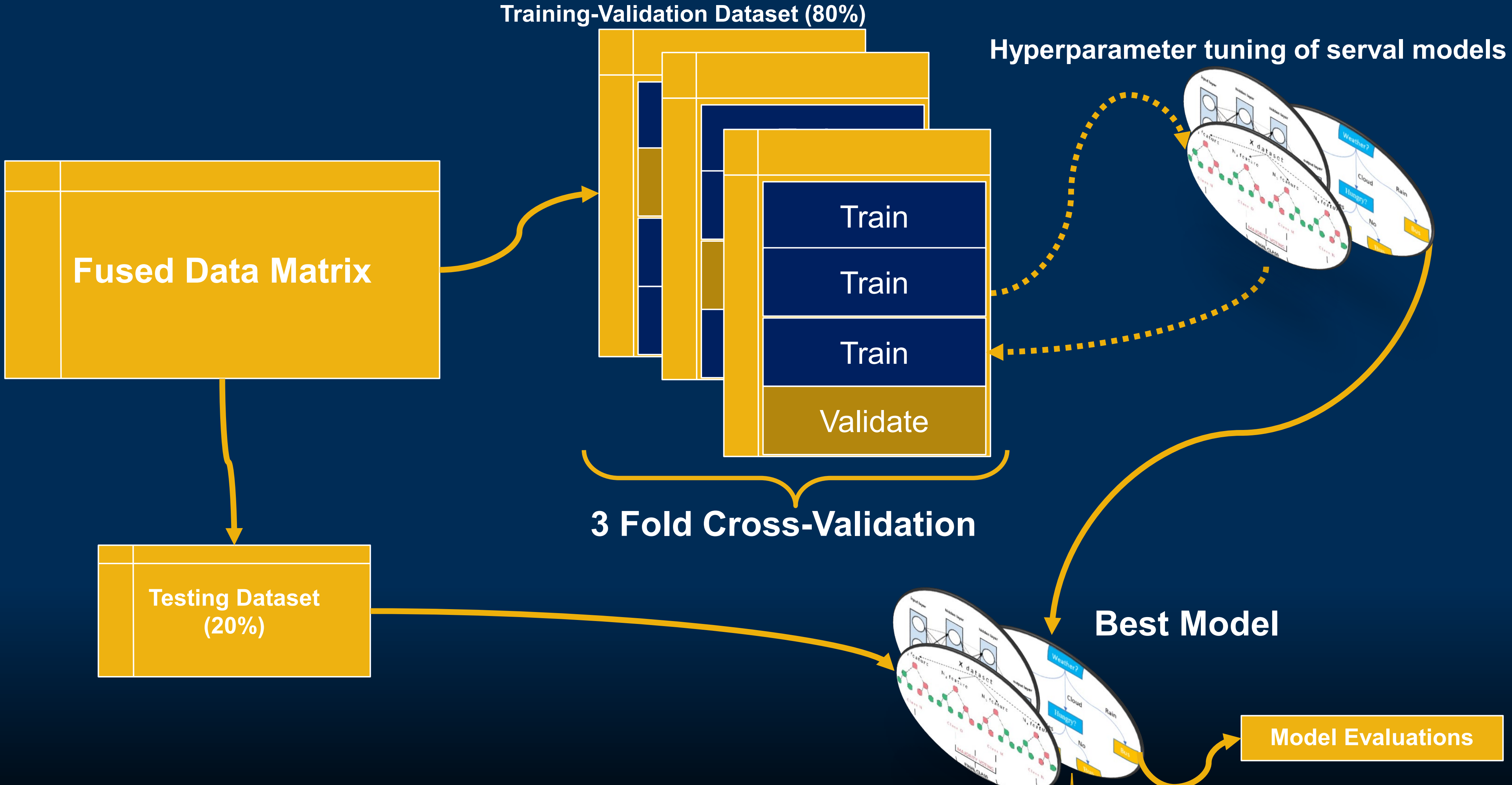
LaGuardia Airport had the most examples



Model Development: ML Algorithms

- 3 popular algorithms are benchmarked for this classification problem
  1. Boosting Ensemble
  2. Random Forests
  3. Multi-Layer Perceptron Neural Network
- Mangortey et al. had shown good results for the first two algorithms [3]
- Methodology is agnostic to the algorithms
  - Other ones could be chosen

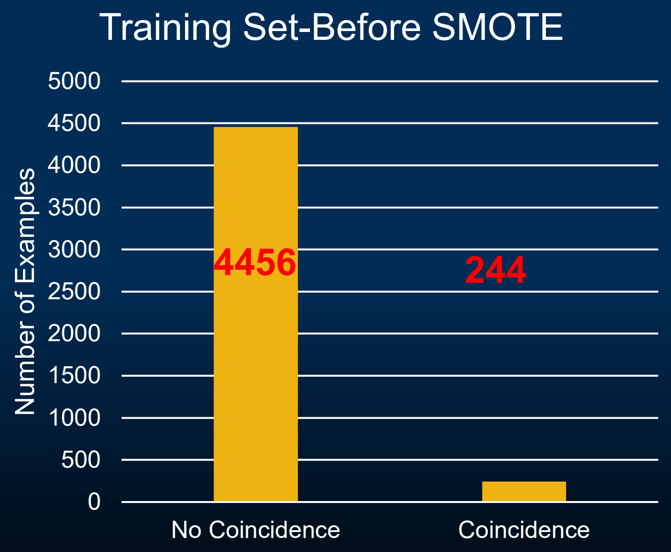
Model Development: Overview of Data Split

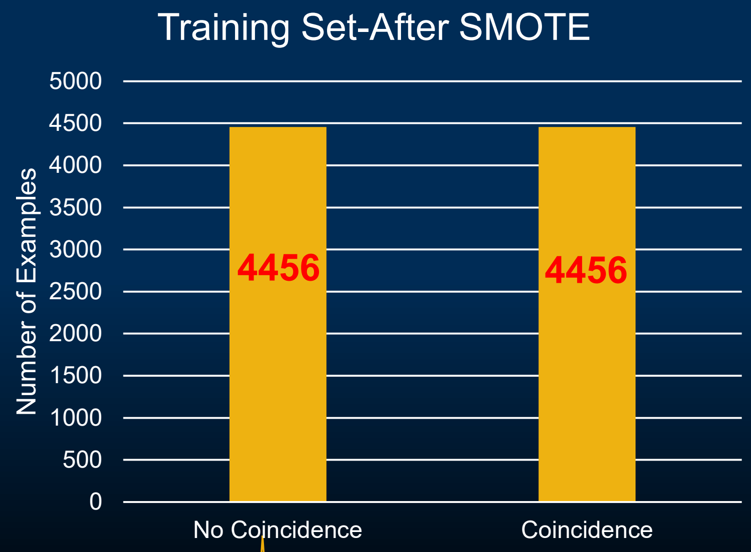


Model Development: Imbalanced Dataset

- Synthetic Minority Over-sampling Technique (SMOTE) was use to deal with class imbalance in the training-validation dataset [16]
  - Uses the k-nearest neighbors of each instance of the minority class to create new points
  - Provides better results than randomly oversampling minority class
  - Inefficient for very large dataset

Ground Delay Program and Ground Stop Coincidence





Motivation

Background

Methodology

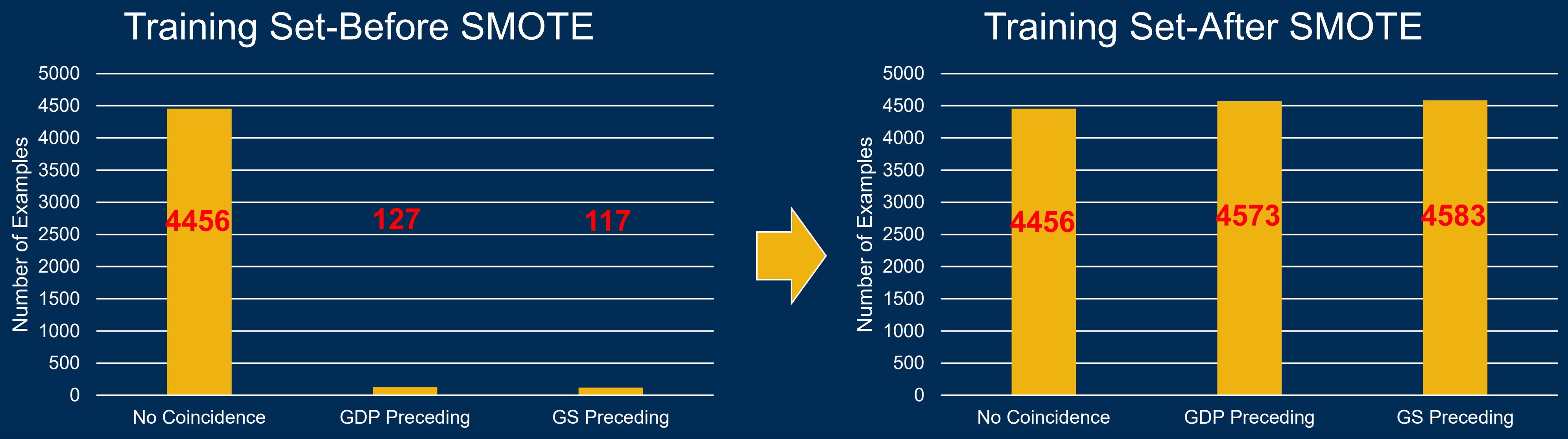
Results

Analysis

Conclusion

Model Development: Imbalanced Dataset

Precedence of Ground Delay Program before Ground Stop, and vise versa, during their coincidence



Motivation

Background

Methodology

Results

Analysis

Conclusion

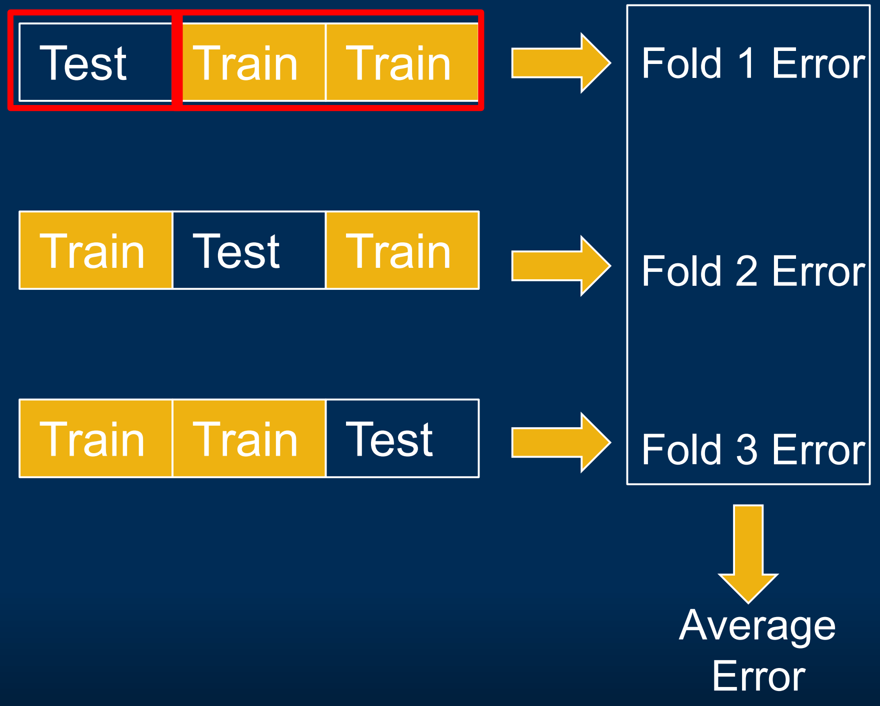
Model Development: Hyperparameter tuning

	Neural Network	Random Forest	Boosting Ensemble
Grid	• Number of hidden layers = [4, 6, 8] • Activation function = [Relu, Elu]	• Max Depth = [30, 50, 70, 110] • Number of estimators = [100, 200, 500, 1000]	• Learning rate = [0.1, 0.001, 0.0001] • Number of estimators = [20, 50, 100, 200]
Number of Combinations	6	16	12

- Each algorithm has multiple hyperparameter settings
  - Grid search allows us to evaluate a given range of settings
- Lower number of combinations for neural network and boosting ensemble
  - Nature of the algorithms make them slow to train
- All combinations (6+16+12=34) were validated using k-fold cross-validation

Model Development: K-Fold Cross-Validation [15]

- K-fold cross-validation can be used to deal with bias-variance trade-off
- Steps of k-fold cross-validation
  - Randomly hold-out a subset of training-validation set
  - Use rest of the set to fit the models
  - Evaluate fitted models on the previously held-out subset
- The steps are repeated k times using different held-out data subset
  - For this work k=3 to reduce computation time
  - Error is averaged across all folds
  - The model with the lowest average was selected as the best model



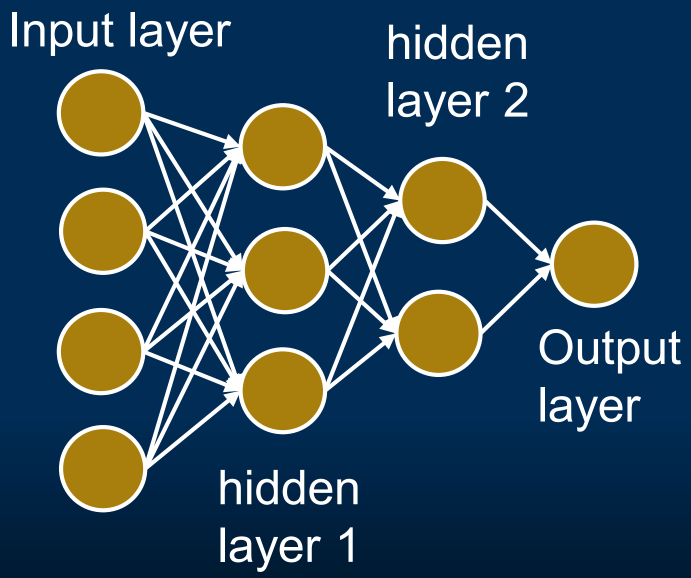
Outline

Motivation → Background → Methodology → **Results** → Analysis → Conclusion

Best Hyperparameters: GS/GDP Coincidence

- Motivation
- Background
- Methodology
- Results
- Analysis
- Conclusion

	Neural Network	Random Forest	Boosting Ensemble
Best Combination	• Activation function: Relu • Number of hidden layers: 8	• Max Depth: 30 • Number of estimator: 100	• Learning rate: 0.1 • Number of estimator: 100



Model Comparison: GS/GDP Coincidence

Motivation
Background
Methodology
Results
Analysis
Conclusion

- We can use the test set and the different metrics to compare the algorithms

Score	Neural Network	Random Forest	Boosting Ensemble
Accuracy	0.967	0.997	0.996
Specificity	0.969	0.999	0.999
Sensitivity	0.929	0.946	0.929
Balanced Accuracy	0.949	0.973	0.964
Kappa Statistics	0.710	0.962	0.952

Random Forest performs better

Model Comparison: GS/GDP Coincidence

Neural Network		
	Actual False	**Actual True**
Predicted False	1084	35
Predicted True	4	52

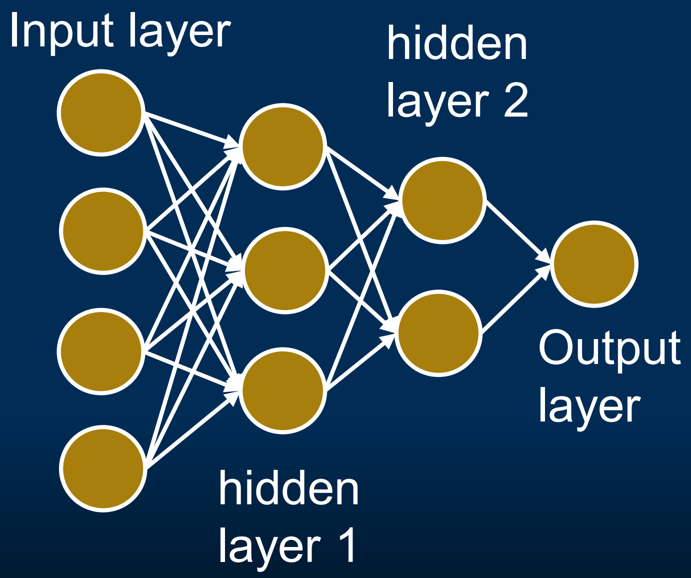
Random Forest		
	Actual False	**Actual True**
Predicted False	1118	1
Predicted True	3	53

Boosting Ensemble		
	Actual False	**Actual True**
Predicted False	1118	1
Predicted True	4	52

- Neural network lower score is due to the higher false positive rate
  - Other algorithms are better at not wrongly predicting a coincidence

Best Hyperparameters: Precedence of GDP before GS and vice versa during Coincidence

	Neural Network	Random Forest	Boosting Ensemble
Best Combination	• Activation function: Relu • Number of hidden layers: 4	• Max Depth: 30 • Number of estimator: 100	• Learning rate: 0.1 • Number of estimator: 50



Model Comparison: Precedence of GDP before GS, and vice versa, during their coincidence

Score	Neural Network	Random Forest	Boosting Ensemble
Accuracy	0.978	0.987	0.986
Balanced Accuracy	0.760	0.833	0.832
Kappa	0.746	0.856	0.841

- Again the Random Forest algorithm performs better
- Neural Network performs similarly as in the previous case
  - More emphasis could put towards the tuning of only this algorithm in the future

Model Comparison: Precedence of GDP before GS, and vice versa, during their coincidence

Neural Network

		Actual		
		Normal	$GDP_{preceding}$	$GS_{preceding}$
Predicted	Normal	1113	4	2
	$GDP_{preceding}$	6	17	5
	$GS_{preceding}$	6	3	19

- Highest false positive than others
- Struggles more to find GDP preceding GS during coincidence
- Highest false negative rate

Random Forest

		Actual		
		Normal	$GDP_{preceding}$	$GS_{preceding}$
Predicted	Normal	1118	0	1
	$GDP_{preceding}$	1	21	6
	$GS_{preceding}$	4	3	21

- Lowest false positive rate
- Able to predict coincidences better but falsely predicts GS preceding GDP during confidence
- Misclassify GS preceding GDP during coincidence by predicting a normal or GDP preceding GS coincidence

Boosting Ensemble

		Actual		
		Normal	$GDP_{preceding}$	$GS_{preceding}$
Predicted	Normal	1116	0	3
	$GDP_{preceding}$	2	20	6
	$GS_{preceding}$	2	4	22

- False positive are misclassification of GS preceding GDP during coincidence
- Also predicts coincidence but misclassify one coincidence type for the other

Outline


Motivation
Background
Methodology
Results
Analysis
Conclusion

Motivation

Background

Methodology

Results

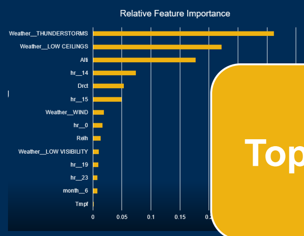
Analysis

Conclusion

Analysis

Analysis 1

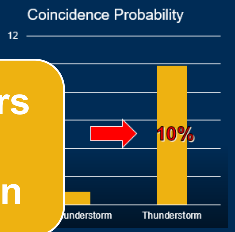
Feature Importance



Top Predictors

- Relative importance plot
- Identify the predictors that are the most important

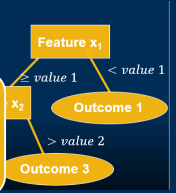
Top Predictors impact quantification



- Partial Dependence Plot
- Quantify how much a change in top predictors can affect predictions

Analysis 2

Surrogate Tree

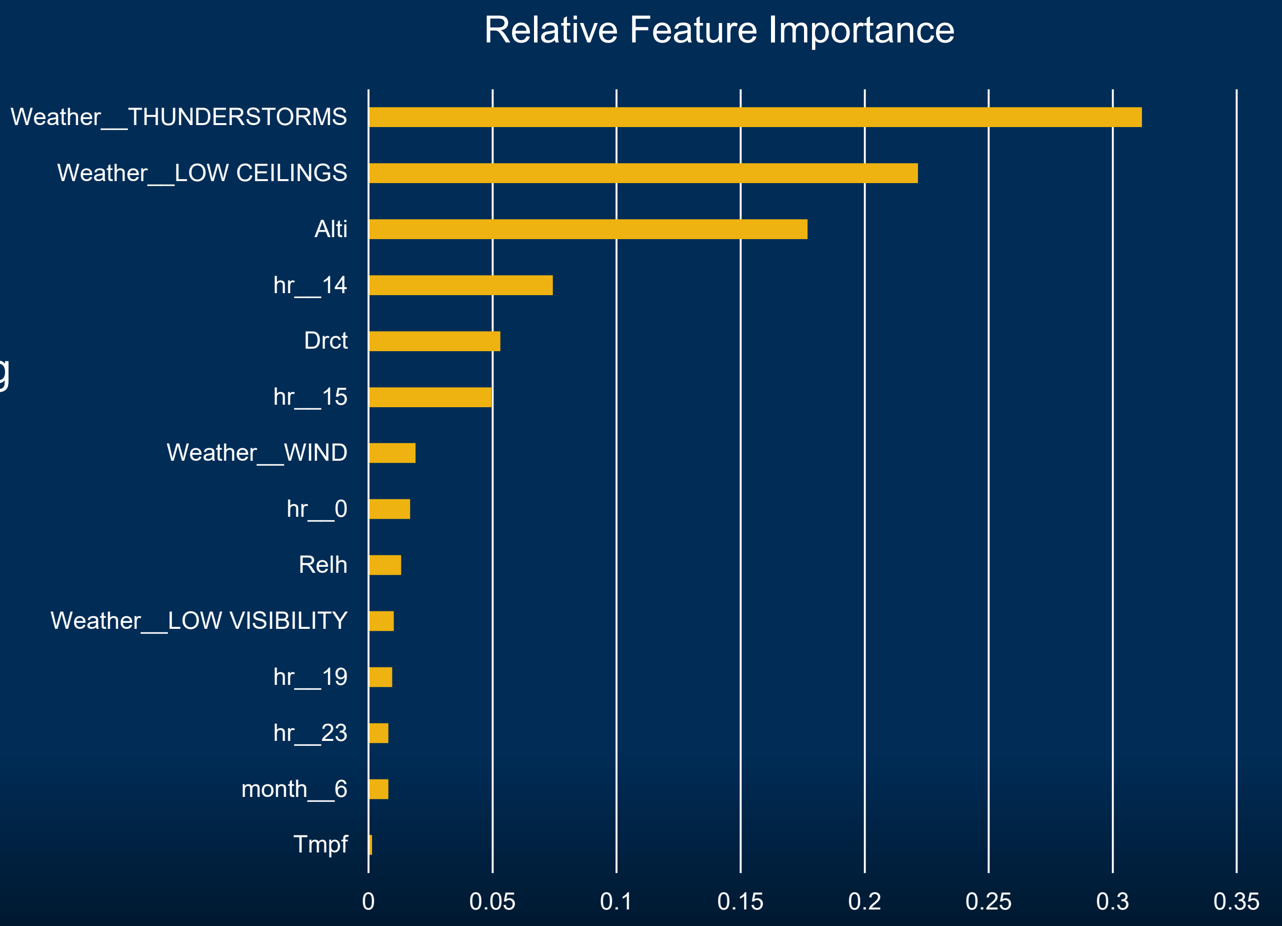


- Decision Tree
- Understand and validate model choices
- Turn a blackbox model into a logical human understandable series of steps

- Analysis 1 and 2 were applied to:
  - GS/GDP coincidence
  - Preceding GDP before GS, and vice versa, during coincidence

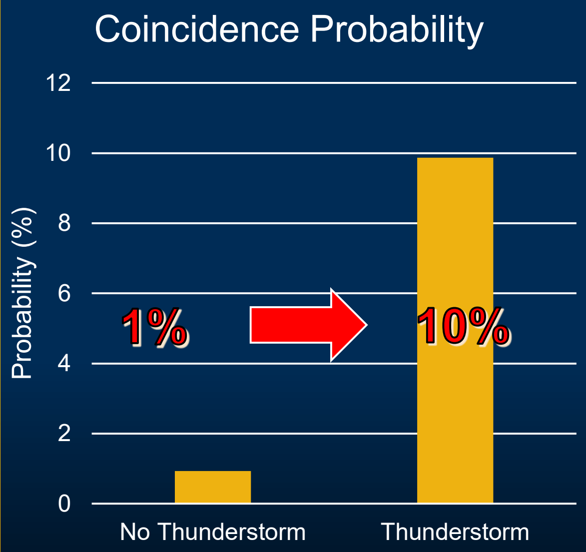
Feature Importance: GS and GDP Coincidence

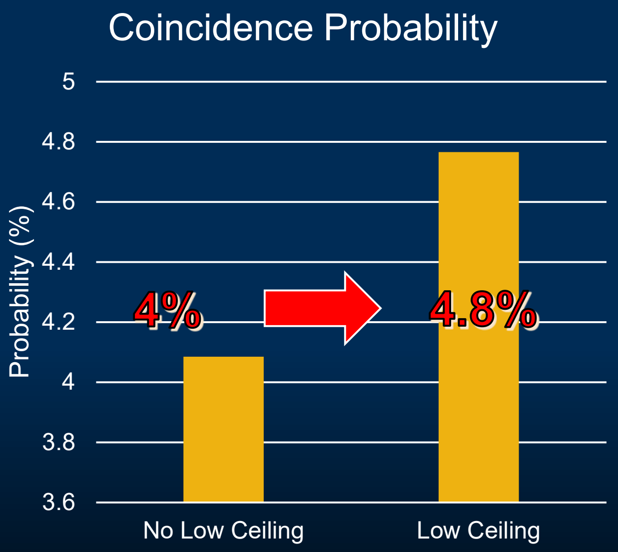
- Best model was the random forest
  - Top 3 most important predictors:
    1. Presence of a thunderstorm
    2. Presence of a low ceiling
    3. Pressure altimeter (inches)

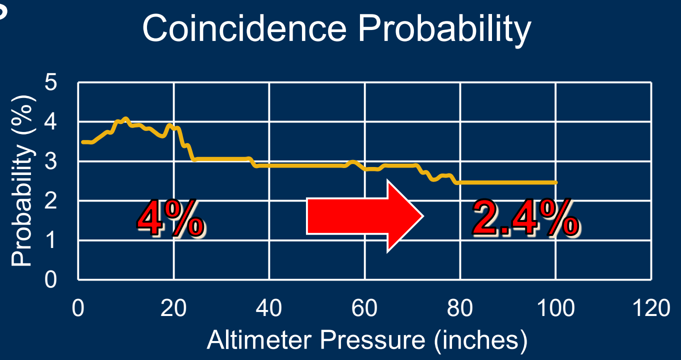


Feature Importance: GS and GDP Coincidence

- Validate the importance of predictors and understand the interaction between them and target feature (coincidence)
  - Use of Partial Dependence Plots (PDP) [9]
  - Shows the dependence between target responses and target features marginalized over the values of other features







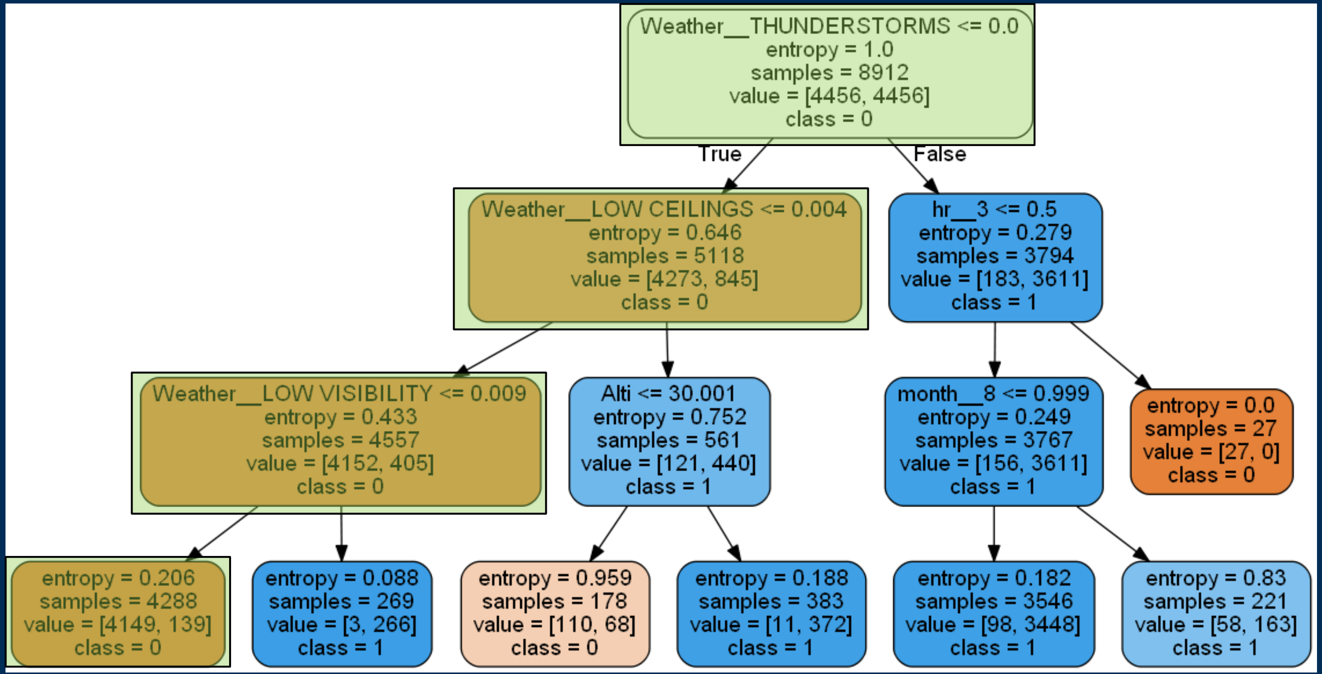
- Changes in model prediction (increase in coincidence probability)
- Overall low probability and low increase
  - Low probability of coincidence
  - More data would allow better quantification of the likelihood change

Surrogate Tree: GS and GDP Coincidence

- Motivation
- Background
- Methodology
- Results
- Analysis
- Conclusion

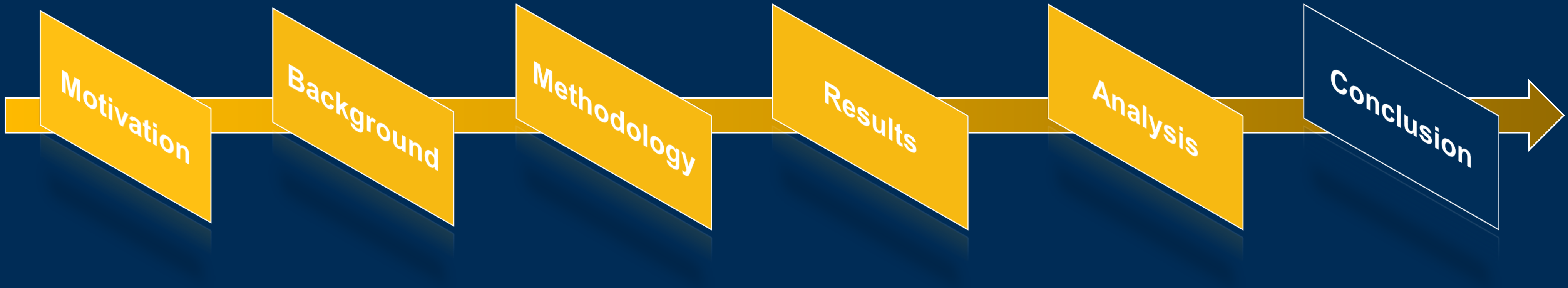
- A global surrogate tree helps make the model more understandable [8]
  - Trains on inputs and predictions of model
  - Approximation of a black-box as a decision tree
- Provides a deterministic way to decide on whether or not to implement multiple TMIs
- Decision of multiple depths were tested
  - Lower depth are more interpretable
  - Higher depth is more accurate but less interpretable

Surrogate Tree: GS and GDP Coincidence



1. There is no thunderstorm
2. There is no low cloud ceiling
3. There is no low visibility
4. Classify as a no coincidence

Outline


Motivation
Background
Methodology
Results
Analysis
Conclusion

Conclusion: Objective 1

Objective 1.1 — Predict the coincidence of weather-related GS and GDP

Objective 1.2 — Predict the precedence of weather-related GDP before GS, and vise versa, during their coincidence

- The Random Forest Model can be used to predict weather-related coincidences
  - Reaching a Kappa statistics of 0.962, and 0.856
  - Traffic Management Personnel could integrate models to current TMI tools for better decision making, planning and training ultimately leading to reduction of delays by improving responses

Conclusion: Objective 2

Objective 2

Analyze prediction models

- Important predictors were discovered (feature importance)
  - Presence of thunderstorm, presence of low ceiling, hour (midnight), and altimeter pressure
- Determined the relations between the important predictors and the target (partial dependence plots)
  - Increase and decrease of coincidence likelihood after changes to predictors
- Learned the logic behind the model decisions (surrogate tree)
  - Streamlining decision making process

Future Work

- Include external data (non-weather related)
  - Allow for better quantification of likelihood of coincidence
  - Better decision trees
- Expand analysis to other airports
- Integrate models to TMI tools
- Presentation of coincidence paper at SciTech

Questions?

References

1. Federal Aviation Administration, June 2019, *Air Traffic By the Numbers*. <https://www.faa.gov/air_traffic/by_the_numbers/media/Air_Traffic_by_the_Numbers_2019.pdf>
2. Transportation Research Board, *Defining and Measuring Aircraft Delay and Airport Capacity Thresholds*. National Academy Press, 2014
3. Mangortey, E., Pinon, O., Puranik, T., and Mavris, D., “Application of Data Fusion and Machine Learning to The Analysis of Ground Delay Programs,” AIAA Aviation Forum 2019
4. Federal Aviation Administration, May 2013, *Traffic Management Initiative Interaction*. http://tfmlearning.faa.gov/TMI_Descriptive_Interaction_Reference_Revised_20140729.pdf
5. Federal Aviation Administration, October 2009, *Traffic Flow Management in the National Airspace System*.< https://www.fly.faa.gov/Products/Training/Traffic_Management_for_Pilots/TFM_in_the_NAS_Booklet_ca10.pdf>
6. “Ground Stop (GS): NBAA - National Business Aviation Association.” *NBAA*, https://nbaa.org/aircraft-operations/airspace/tfm/tools-used-for-traffic-flow-management/ground-stop-gs/
7. Murphy, Kevin P. *Machine Learning: a Probabilistic Perspective*. Cram101, 2013.

References

8. Christopher Molnar. *A Guide for Making Black Box Models Explainable*, 2019, https://christophm.github.io/interpretable-ml-book/global.html
9. Scikit-learn. *Partial Dependence Plots*. https://scikit-learn.org/stable/modules/partial_dependence.html
10. https://www.metabunk.org/fire-at-faa-facility-closes-chicago-airspace.t4575/
11. Federal Aviation Administration, "Air Traffic By the Numbers," , June 2019. URL https://www.faa.gov/air_traffic/by_the_numbers/media/Air_Traffic_by_the_Numbers_2019.pdf.
12. Jixin, L., "Optimizing Key Parameters of Ground Delay Program with Uncertain Airport Capacity," *Journal of Advanced Transportation*, Vol. 2017, No. 6, 2017, p. 19. doi:10.1155/2017/7494213, special issue on Transportation Simulation Advances in Air Transportation Research

References

13. Wang, Y., “ANALYSIS AND PREDICTION OF WEATHER IMPACTED GROUND STOP OPERATIONS,”*33*rdDigitalAvionicsSystemsConference; 2014
14. Smith, D. A., “DECISION SUPPORT TOOL FOR PREDICTING AIRCRAFT ARRIVAL RATES FROM WEATHER FORECASTS,” Ph.D. thesis, Georgia Mason University, 2008.
15. Song, Lee. *Overfitting and Cross-Validation.* August 2019. PowerPoint Presentation
16. https://www.cs.cmu.edu/afs/cs/project/jair/pub/volume16/chawla02a-html/node6.html
17. https://nbaa.org/aircraft-operations/airspace/tfm/tools-used-for-traffic-flow-management/