CLASSIFICATION OF AN AIRCRAFT'S LATERAL DEVIATION

Oscar Cardec

PROJECT OVERVIEW

PROJECT BACKGROUND

DATASET, KPI & FINDINGS

PREDICTIVE MODELS

RESULTS & RECOMMENDATIONS

PROJECT DESCRIPTION

PROJECT DESCRIPTION

- ✓ Stakeholder: Federal Aviation Administration (FAA)
- ✓ Evaluation of aircraft lateral deviations from the permissible route
- ✓ 2008 Dataset ~10 Million observations across the CONUS 20 Air Route Traffic Control Centers (ARTCC)
- ✓ Background on prior analytics efforts
  - Paglione, et al, 2009 - 8th USA/Europe Air Traffic Management Research and Development Seminar
  - Nelson & Paglione, 2013 – American Institute of Aeronautics and Astronautics

PROJECT STATEMENT

To analyze the proposed dataset, employ new statistical methods, improve predictive analytics, and ultimately develop a novel machine learning algorithm to enhance the near-real-time lateral anomaly detection on a given flight.

PROJECT SCOPE

Importance

- Accurate prediction of an aircraft position is vital
- Traffic management and safety responsibility
- Impact across the airline industry operations & logistics

Focus

- Identification of correlation across the given variables
- Systematic approach to differentiate between variance and deviancy
- Improve FAA's decision support tools (DSTs) in mitigating workload constrains

DATASET DESCRIPTION

DATA SET DESCRIPTION

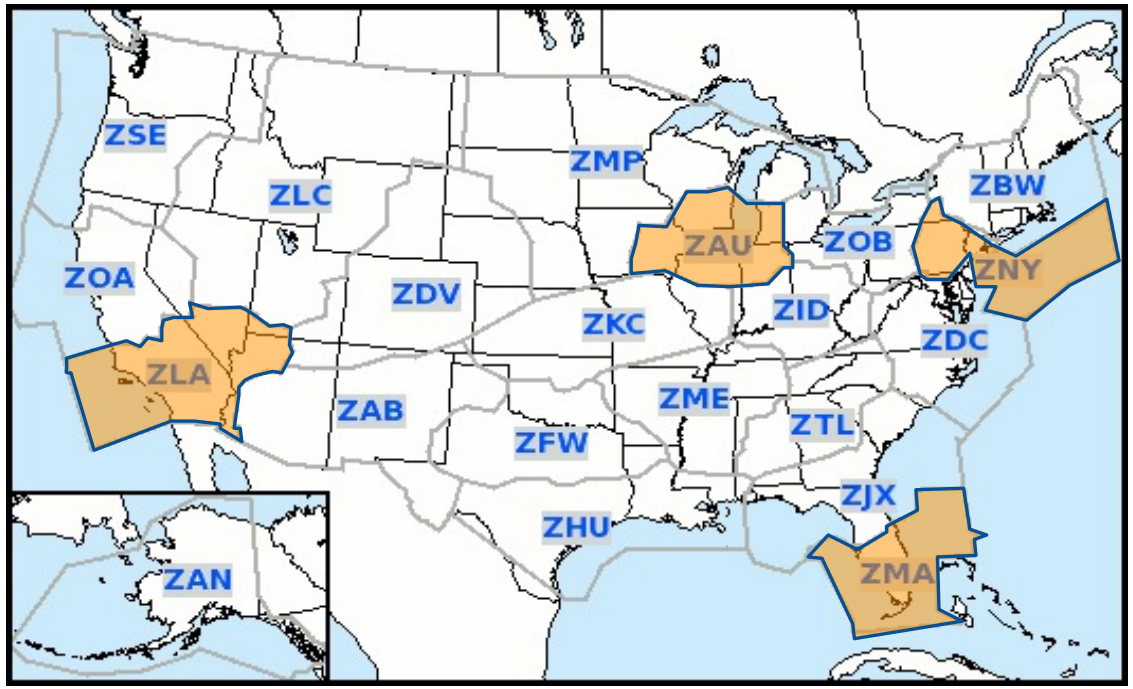
Original Data from 20 ARTCC

- ~ 500,000 observations per ARTCC (Total 10 million)

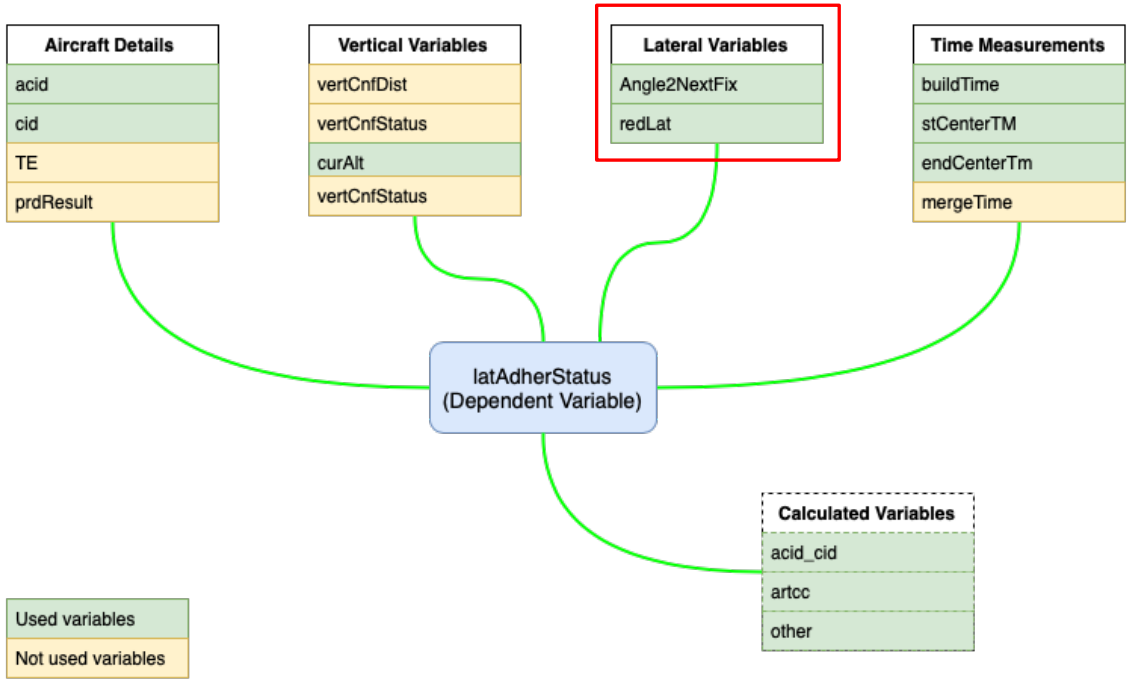
Sampled population

- 4 ARTCC
  - ZAU 549,977
  - ZLA 547,118
  - ZMA 473,121
  - ZNY 417,852

  Total 1,988,068

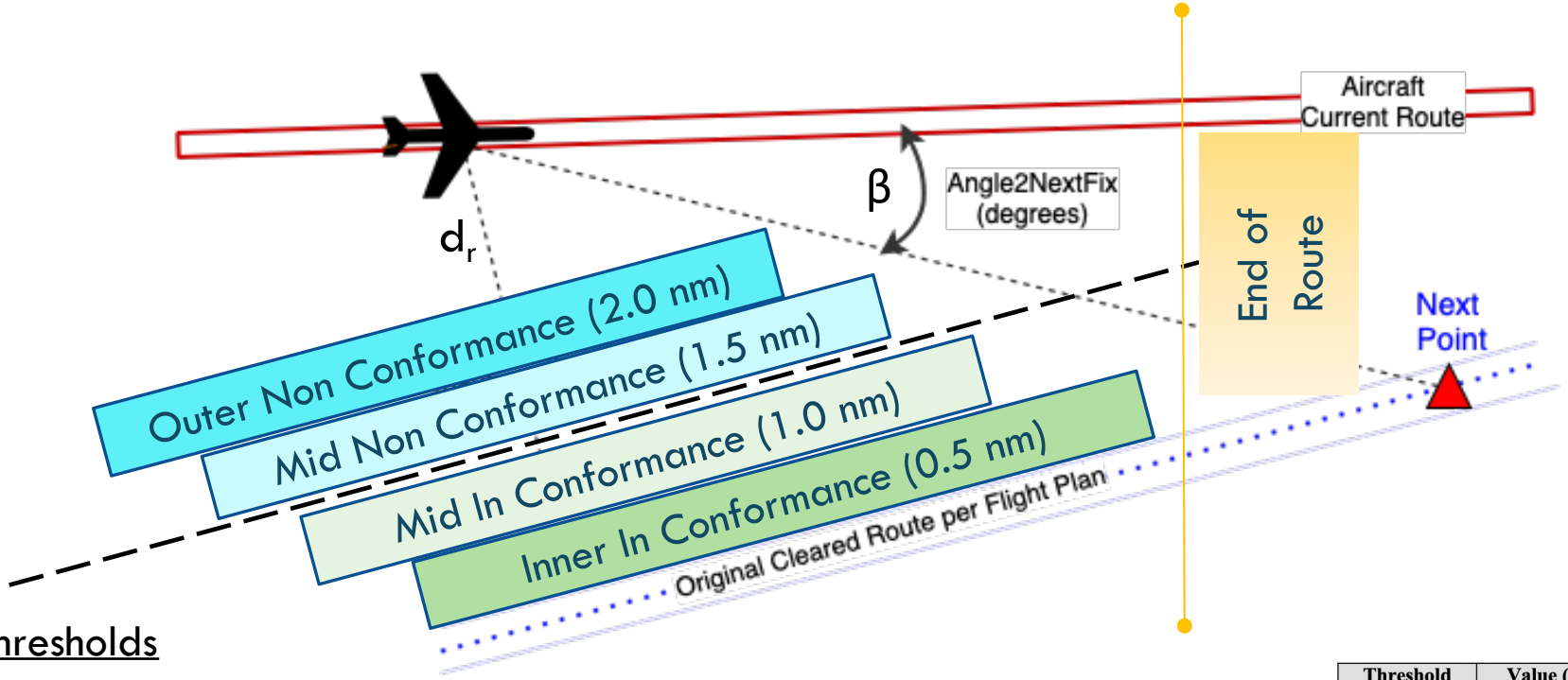


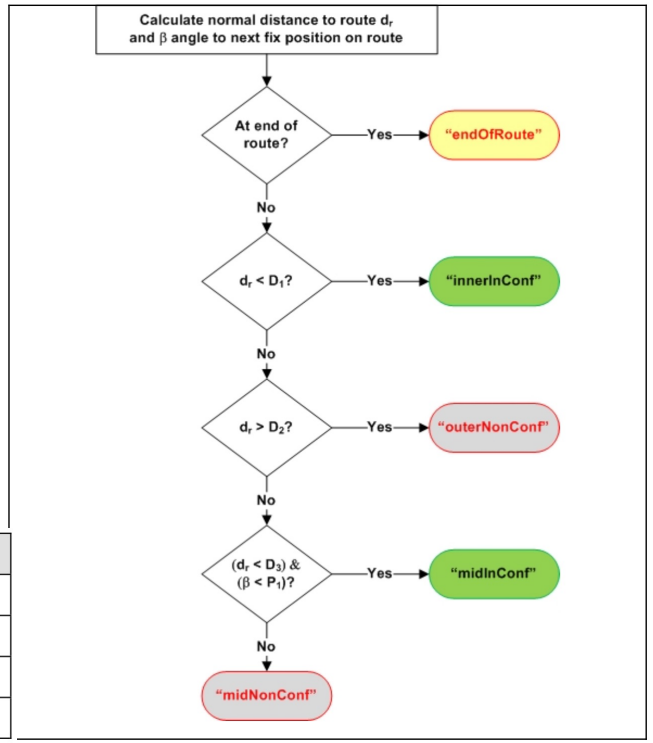
VARIABLES ON THE DATASET



#	Field Name	Description	Applicable	Data Type
1	acid	Aircraft ID for flight	Yes	String
2	cid	Computer ID for flight	Yes	String
3	buildTime	Time in seconds of day that measurement was taken	Yes	Integer
4	TE	Trajectory Engine Code	No	String
5	curAlt	Current altitude of position report	No	Integer
6	stCenterTm	Center start time in seconds of day that aircraft in control within Center	Yes	Integer
7	endCenterTm	Center end time in seconds of day that aircraft in control within Center	Yes	Integer
8	mergeTime	Algorithm specific of trajectory engine	No	Integer
9	vertCnfStatus	Vertical conformance status	No	String
10	vertCnfDist	Vertical conformance distance in feet	No	Integer
11	redLat	Lateral deviation of flight in nautical miles	Yes	Real
12	Angle2NextFix	Angle to next fix on flight plan from current position	Yes	Real
13	latAdherStatus	Lateral conformance status code	Yes	String
14	prdResult	Algorithm specific of trajectory engine	No	Real

VARIABLES ON THE DATASET – CONT.





Thresholds

Threshold	Value (units)
D_1	0.5 (nm)
D_2	1.5 (nm)
D_3	1.0 (nm)
P_1	30 (deg)

Image source: Paglione, et al. 2009)

KEY PERFORMANCE INDICATORS

- Sampled distribution conforms to the central limit theorem and sampled data is representative of the entire population
- Balanced classes across the measured target variable
- Establish a baseline, and improve classification models accuracy

PREPARATION REQUIREMENTS

Preparation Steps

- Selection of relevant variables, and balancing of classes
- Feature engineering (concatenate, rename, transform, handling of outliers, normalization, etc.)
- Dimensionality expansion ($a^2 + b^2 = c^2$) $c = b / \sin B$
- Additional features (moving average, absolute weighted mov average, fix point distance)

Preparation Tools

- RStudio
- Jamovi

ANALYSIS STRATEGY

- Business problem/data understanding
- Extraction & transformation
- Assessment of descriptive statistics
- Data preprocessing
- Visualizations and preliminary findings
- Predictive Models & Evaluation
- Analysis of Results

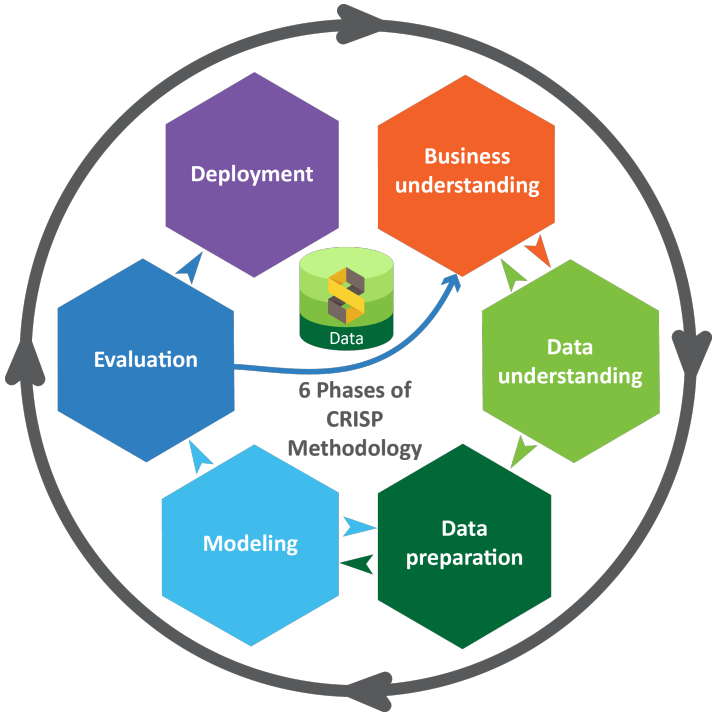
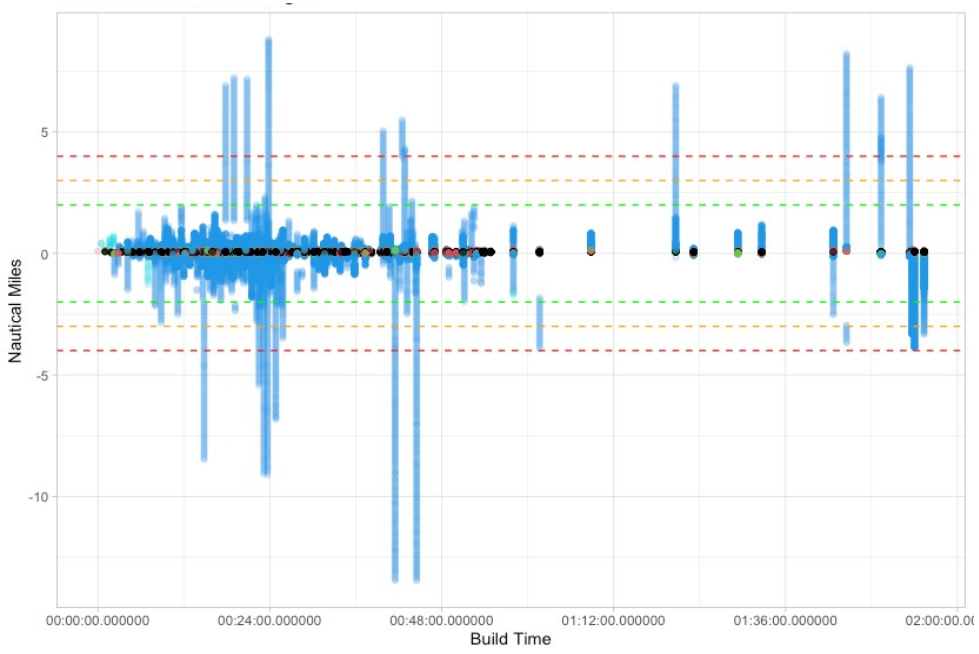


Image Source

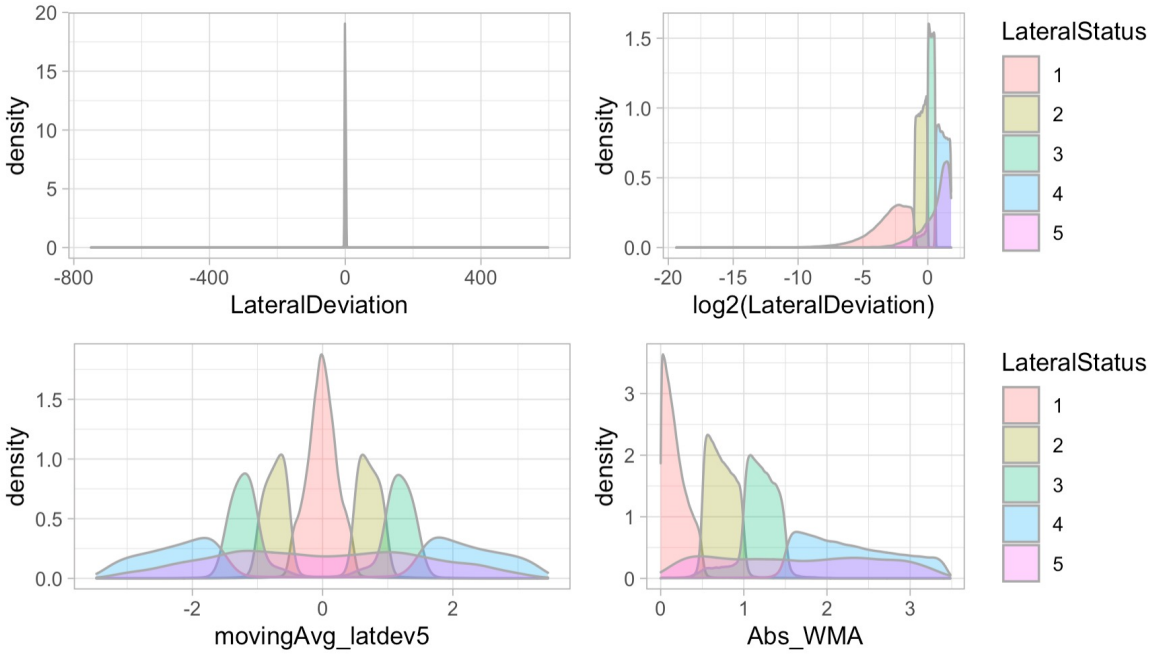
DATASET INSIGHTS & FINDINGS

DATASET INSIGHTS & FINDINGS

Deviations While in the Zone

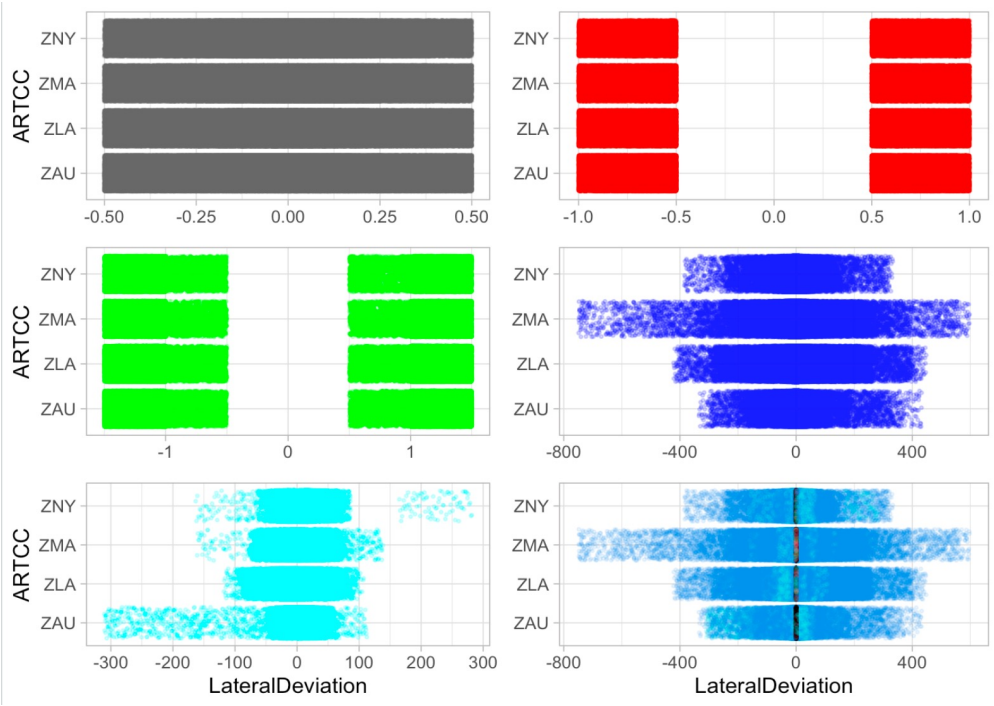


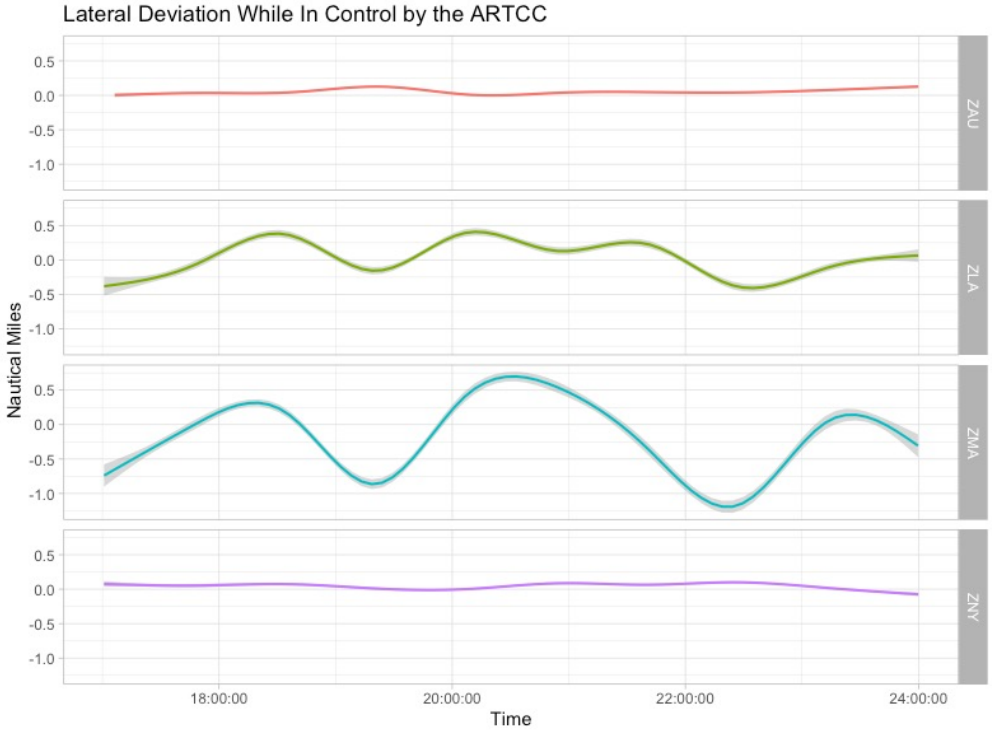
Granular View of Lateral Deviation



DATASET INSIGHTS & FINDINGS

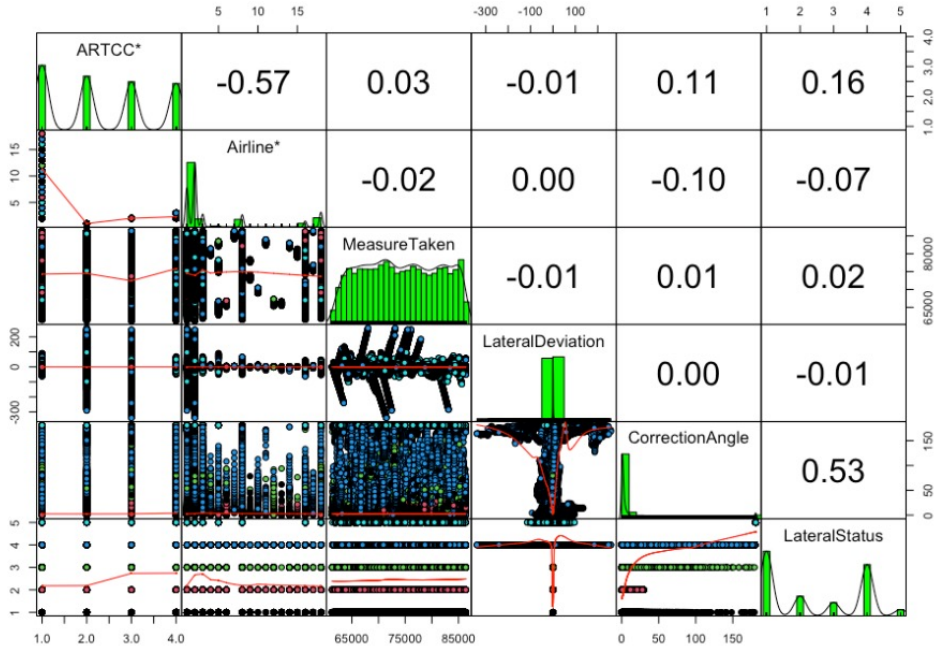
Lateral Deviation per ARTCC



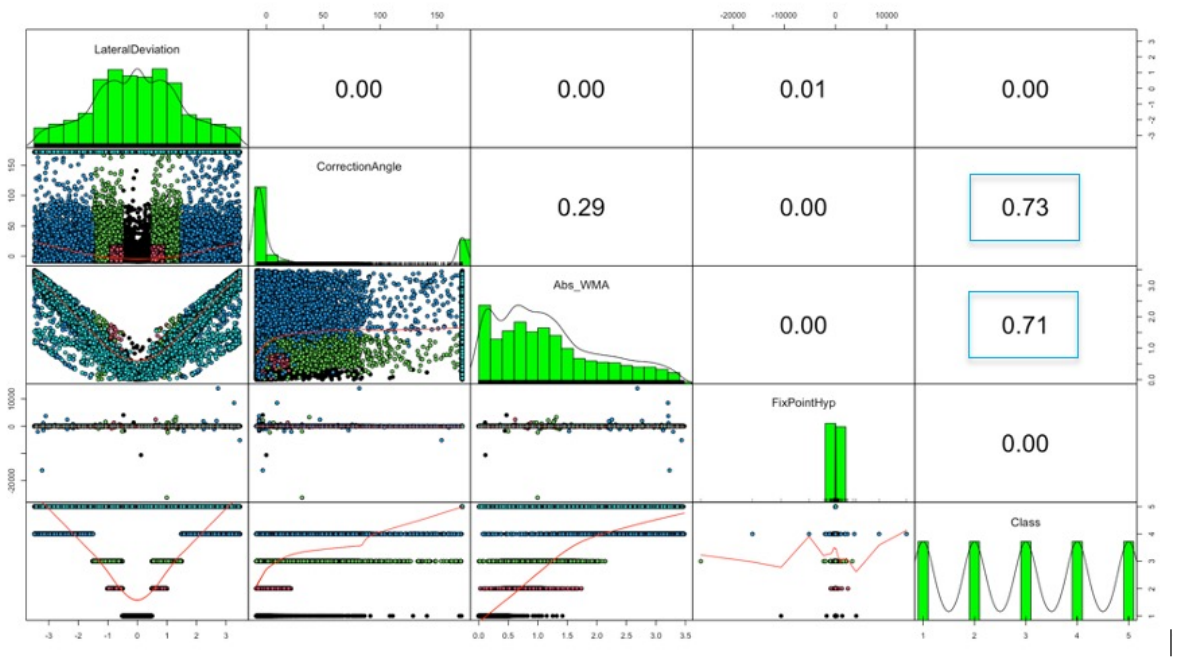


DATASET INSIGHTS & FINDINGS

Original Data

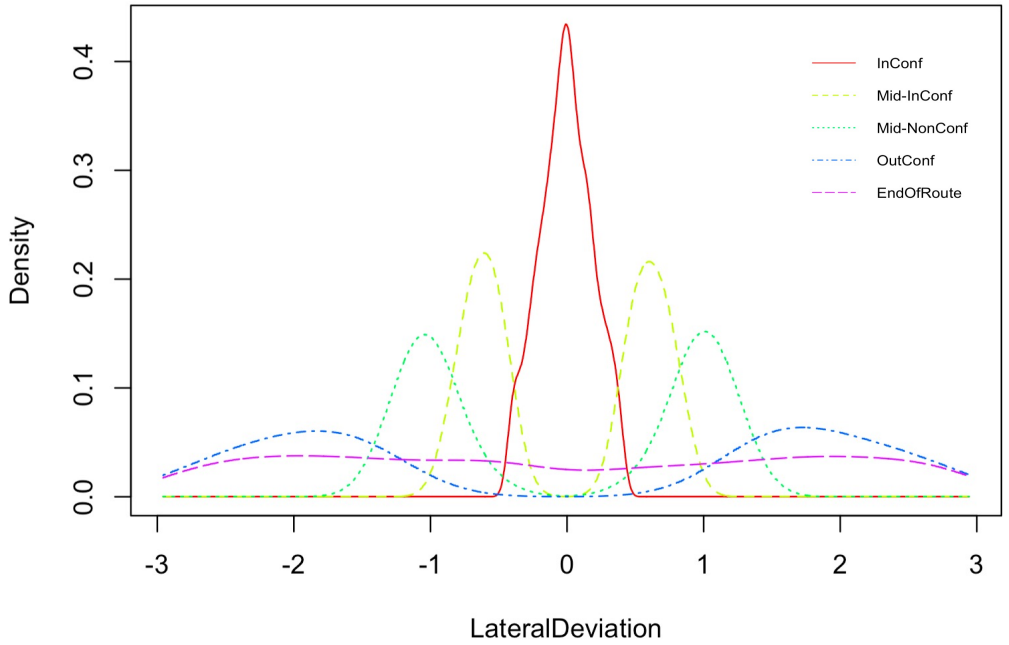


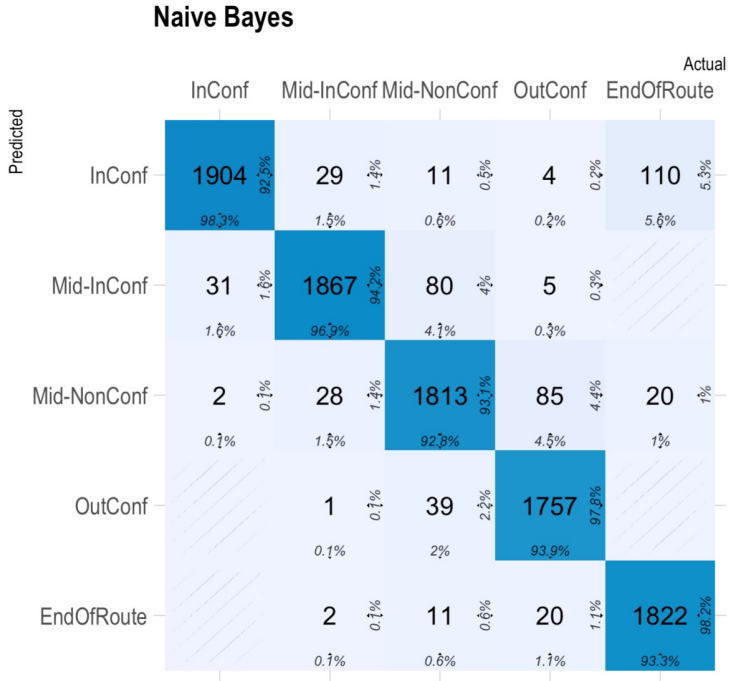
Transformed Data



PREDICTIVE MODELS

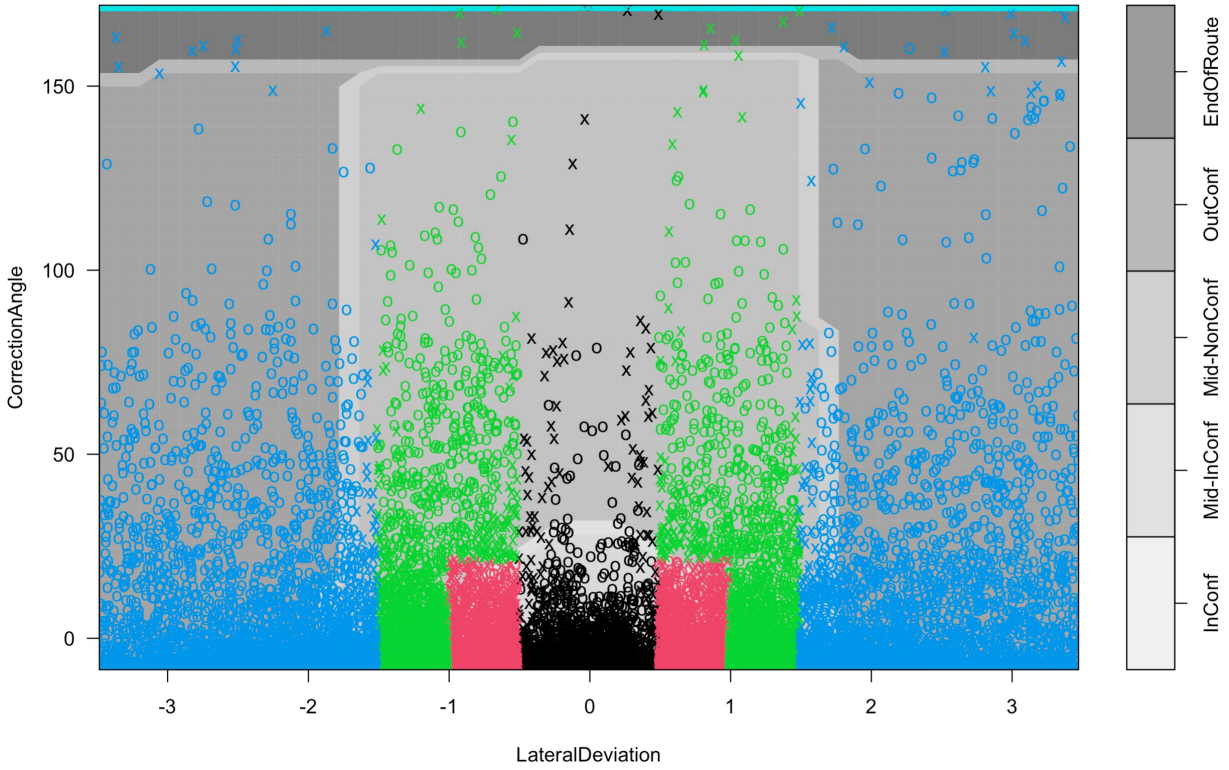
NAÏVE BAYES



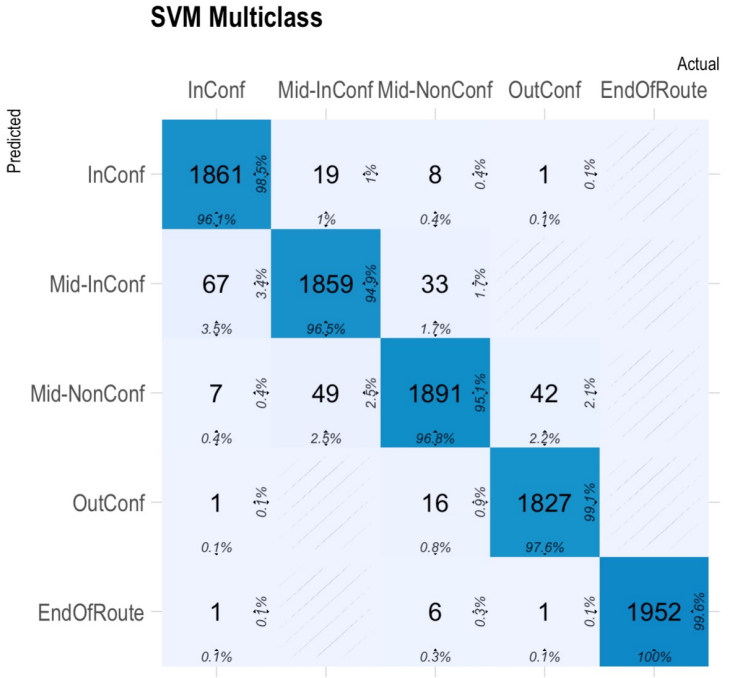


Model	Dataset	Accuracy	Recall	Specificity	Precision	F1 Score	Kappa	Runtime (seconds)
Naïve Bayes	Training	0.9488	0.9488	0.9872	0.9499	0.9488	0.9360	1.60
	Cross-validate	0.9504	0.9504	0.9875	0.9514	0.9505	0.9381	

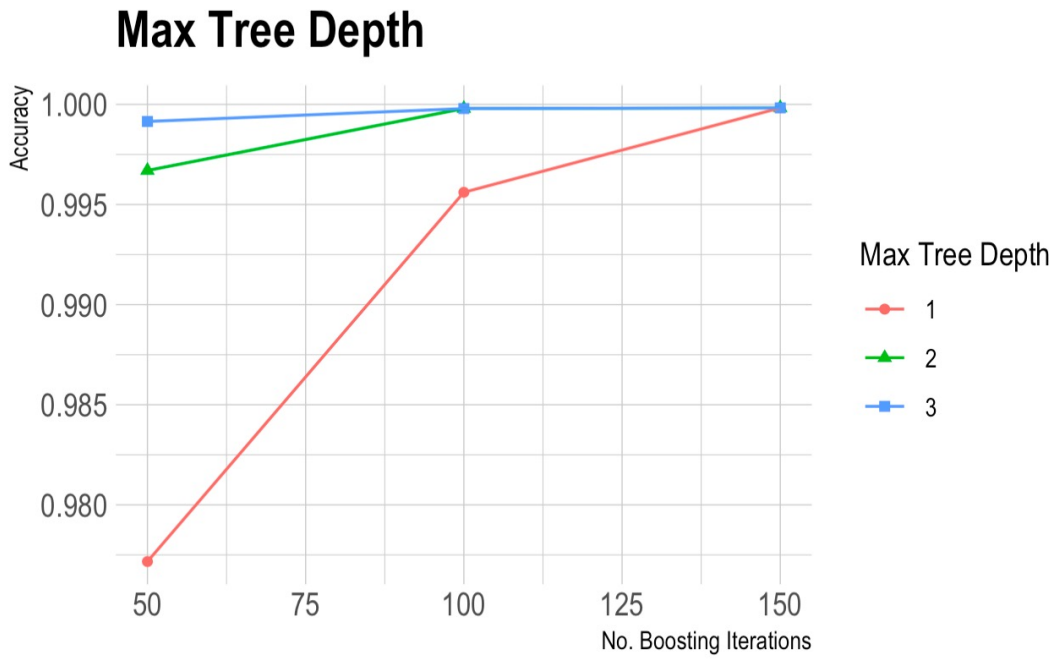
SUPPORT VECTOR MACHINE

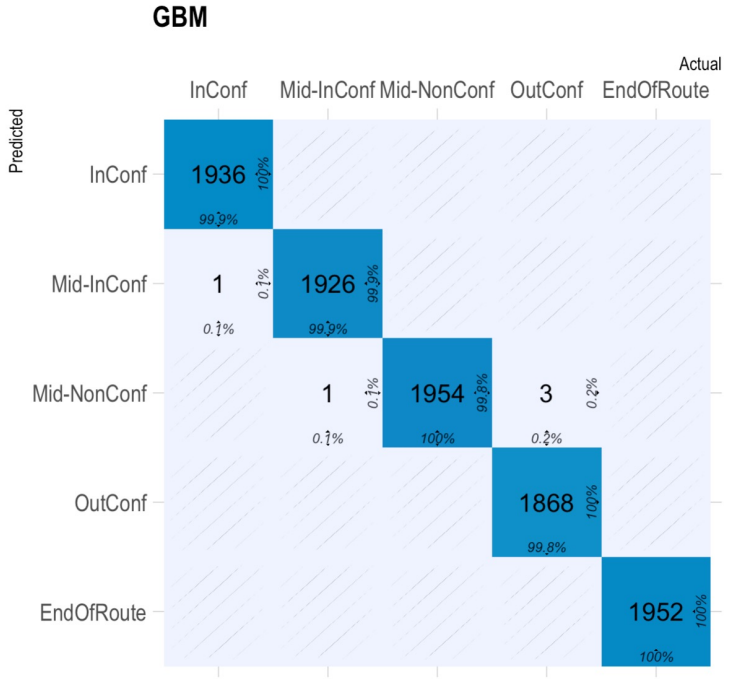


Model	Dataset	Accuracy	Recall	Specificity	Precision	F1 Score	Kappa	Runtime (seconds)
SVM-M	Training	0.9721	0.9720	0.9930	0.9722	0.9721	0.9651	5.29
	Cross-validate	0.9739	0.9739	0.9934	0.9743	0.9740	0.9675	



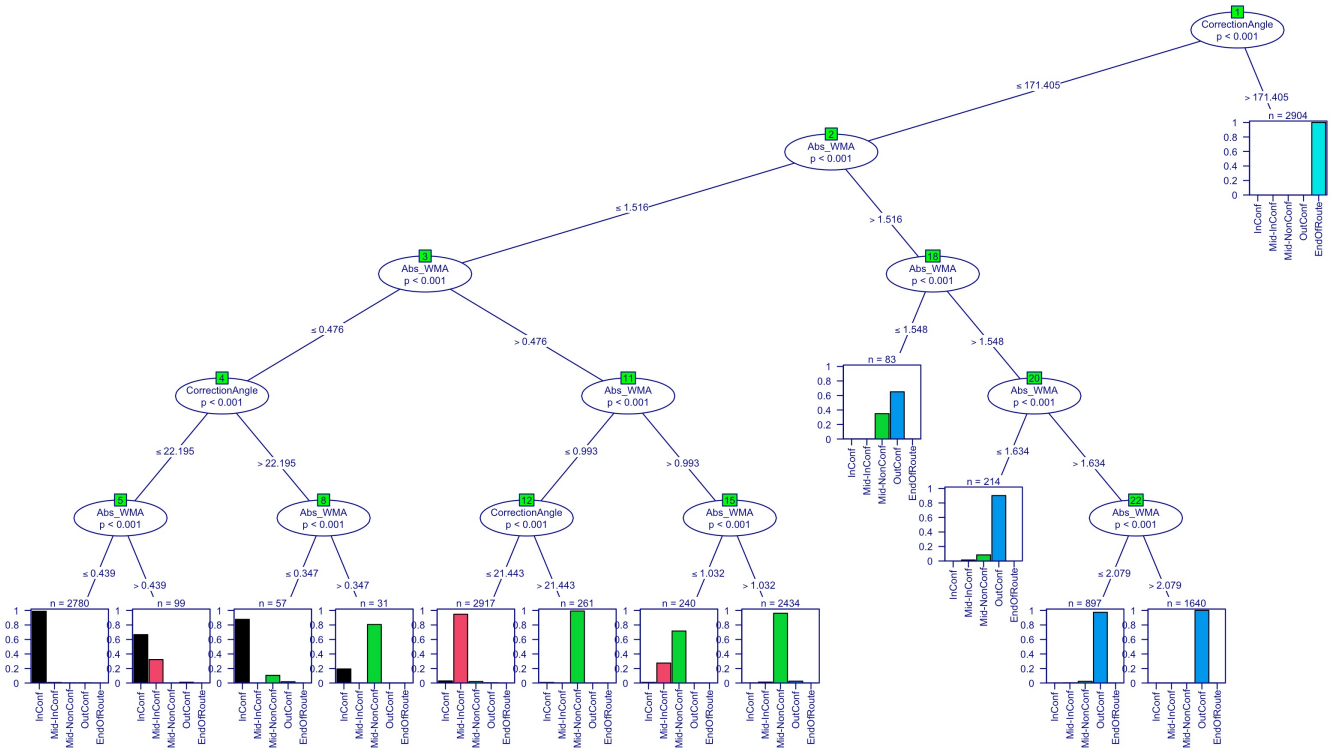
GRADIENT BOOSTING MACHINE



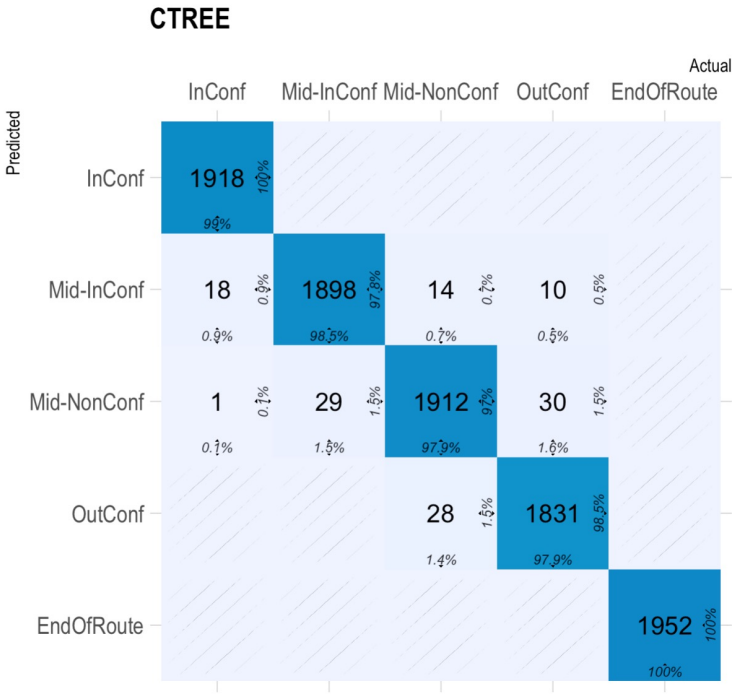


Model	Dataset	Accuracy	Recall	Specificity	Precision	F1 Score	Kappa	Runtime (seconds)
GBM	Training	0.9997	0.9997	0.9999	0.9997	0.9997	0.9997	43.99
	Cross-validate	0.9994	0.9994	0.9998	0.9994	0.9994	0.9993	

CONDITIONAL INFERENCE TREE



Model	Dataset	Accuracy	Recall	Specificity	Precision	F1 Score	Kappa	Runtime (seconds)
CTREE	Training	0.9851	0.9851	0.9962	0.9851	0.9851	0.9813	7.63
	Cross-validate	0.9865	0.9864	0.9966	0.9865	0.9864	0.9831	

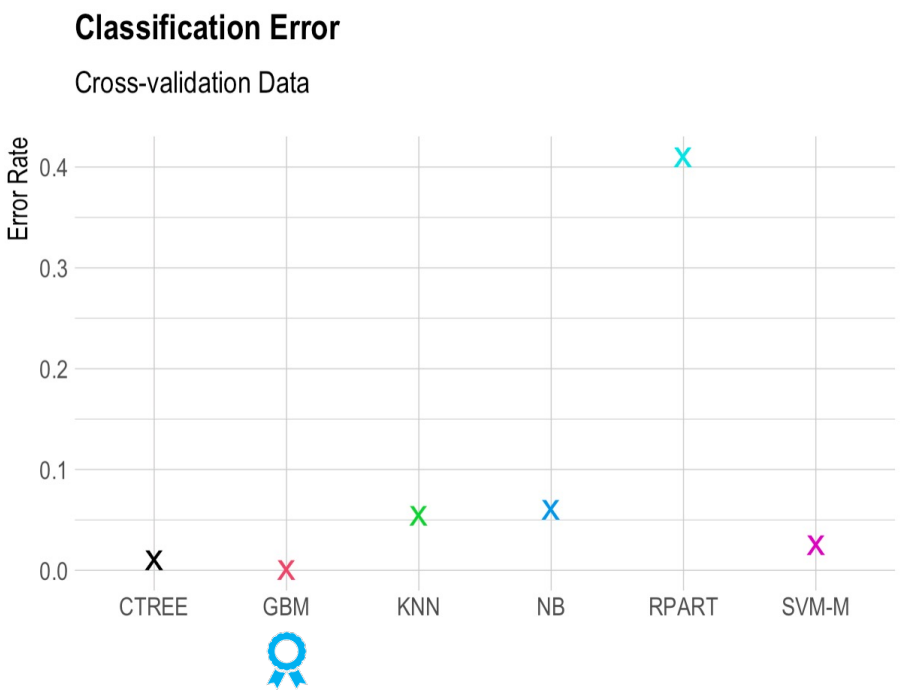


RESULTS AND RECOMMENDATIONS

CLASSIFICATION OUTCOME

Model	Dataset	Accuracy	Recall	Specificity	Precision	F1 Score	Kappa	Runtime (seconds)
Naïve Bayes	Training	0.9488	0.9488	0.9872	0.9499	0.9488	0.9360	1.60
	Cross-validate	0.9504	0.9504	0.9875	0.9514	0.9505	0.9381	
SVM-M	Training	0.9721	0.9720	0.9930	0.9722	0.9721	0.9651	5.29
	Cross-validate	0.9739	0.9739	0.9934	0.9743	0.9740	0.9675	
CTREE	Training	0.9851	0.9851	0.9962	0.9851	0.9851	0.9813	7.63
	Cross-validate	0.9865	0.9864	0.9966	0.9865	0.9864	0.9831	
GBM	Training	0.9997	0.9997	0.9999	0.9997	0.9997	0.9997	43.99
	Cross-validate	0.9994	0.9994	0.9998	0.9994	0.9994	0.9993	
RPART	Training	0.5914	0.5914	0.8983	NaN	NaN	0.4473	2.11
	Cross-validate	0.5892	0.5926	0.8973	NaN	NaN	0.4480	
KNN	Training	0.9404	0.9403	0.9851	0.9415	0.9406	0.9257	10.50
	Cross-validate	0.9437	0.9437	0.9859	0.9446	0.9439	0.9298	

Baseline Runner-up Champion



NOTES & RECOMMENDATIONS

- Near real time (NRT) detection system capability
- Early anticipation of congested hubs or regions
- Data preprocessing is critical for this kind of analysis
- Further unsupervised learning research is highly advised
- Recommending to use updated data with additional attributes (e.g., Lat/Long, Weather, GPS, etc.)
- Elevate research to address an actual FAA operational need (e.g., deploy & test the model)

REFERENCES

Abbott, D. (2014). *Advance predictive analytics: Principles and techniques for the professional data analyst*. Wiley.

Brownlee, J. (2020). A Gentle Introduction to the Gradient Boosting Algorithm for Machine Learning. https://machinelearningmastery.com/gentle-introduction-gradient-boosting-algorithm-machine-learning/

Bruce, P., & Bruce, A. (2017). *Practical statistics for data science: 50 Essential concepts*. O'Reilly.

Elwell, D. K. (2019). FAA Strategic Plan FY 2019-2022. https://www.faa.gov/about/plans_reports/media/FAA_Strategic_Plan_Final_FY2019-2022.pdf

EMC Edu Services. (2015). *Data science and big data analytics. Discovering, analyzing, visualizing, and presenting data.* Wiley.

FAA. (2021) Aeronautical Information Manual. Section 3. En-Route Procedures. https://www.faa.gov/air_traffic/publications/atpubs/aim_html/chap5_section_3.html

(2021) Air Traffic Control, Order JO 7110.65Z. U.S. Department of Transportation, Federal Aviation Administration, Air Traffic Organization Policy. https://www.faa.gov/documentLibrary/media/Order/7110.65Z_ATC_Bsc_dtd_6-17-21.pdf

(2021). The FAA Mission and Vision. https://www.faa.gov/about/mission/

Friedman, J. H. (1999). Stochastic Gradient Boosting. https://statweb.stanford.edu/~jhf/ftp/stobst.pdf

Majka, M (2019). *naivebayes: High Performance Implementation of the Naive Bayes Algorithm in R*. R package version 0.9.7, https://CRAN.R-project.org/package=naivebayes.

Paglione, M., Bayraktutar, I., McDonald, G., and Bronsvoort, J. (2009). Lateral Intent Error's Impact on Aircraft Prediction. Air Traffic Control Quarterly 2010 18:1, 29-62.

Ravanshad, A. (2018). Gradient Boosting vs. Random Forest. https://medium.com/@aravanshad/gradient-boosting-versus-random-forest-cfa3fa8f0d80

Stats.libretext.org (2021) The Empirical Rule and Chebyshev's Theorem. https://stats.libretexts.org/...The_Empirical_Rule_and_Chebyshev%27s_Theorem

Widmann, M. (2020). Cohen's Kappa: What It Is, When to Use It, and How to Avoid Its Pitfalls. https://thenewstack.io/cohens-kappa-what-it-is-when-to-use-it-and-how-to-avoid-its-pitfalls/

QUESTIONS