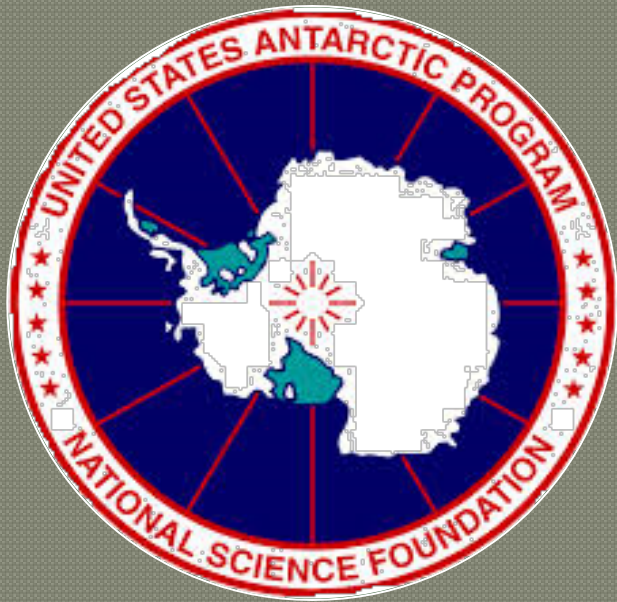


# Applying a Random Forest Classification Model for Fog at Pegasus Airfield

Joey Snarski

Research Analyst  
Scientific Research Corporation





# Overview

- Background
- Data
- Results
- Future Directions





# Decision Trees and Random Forests



- Random Forests = ensembles of decision trees
- Decision trees “vote” for a class



# Observational Data

- 11411 manual observations
  - Dec 2009 – Jan 2017
  - 7 predictor variables
  - No special observations
- Partial, shallow, freezing, mist = fog
- Local to the aerodrome
- No cloud ceiling → 25kft
- Other missing values were imputed using miniature random forests



# Initial Run

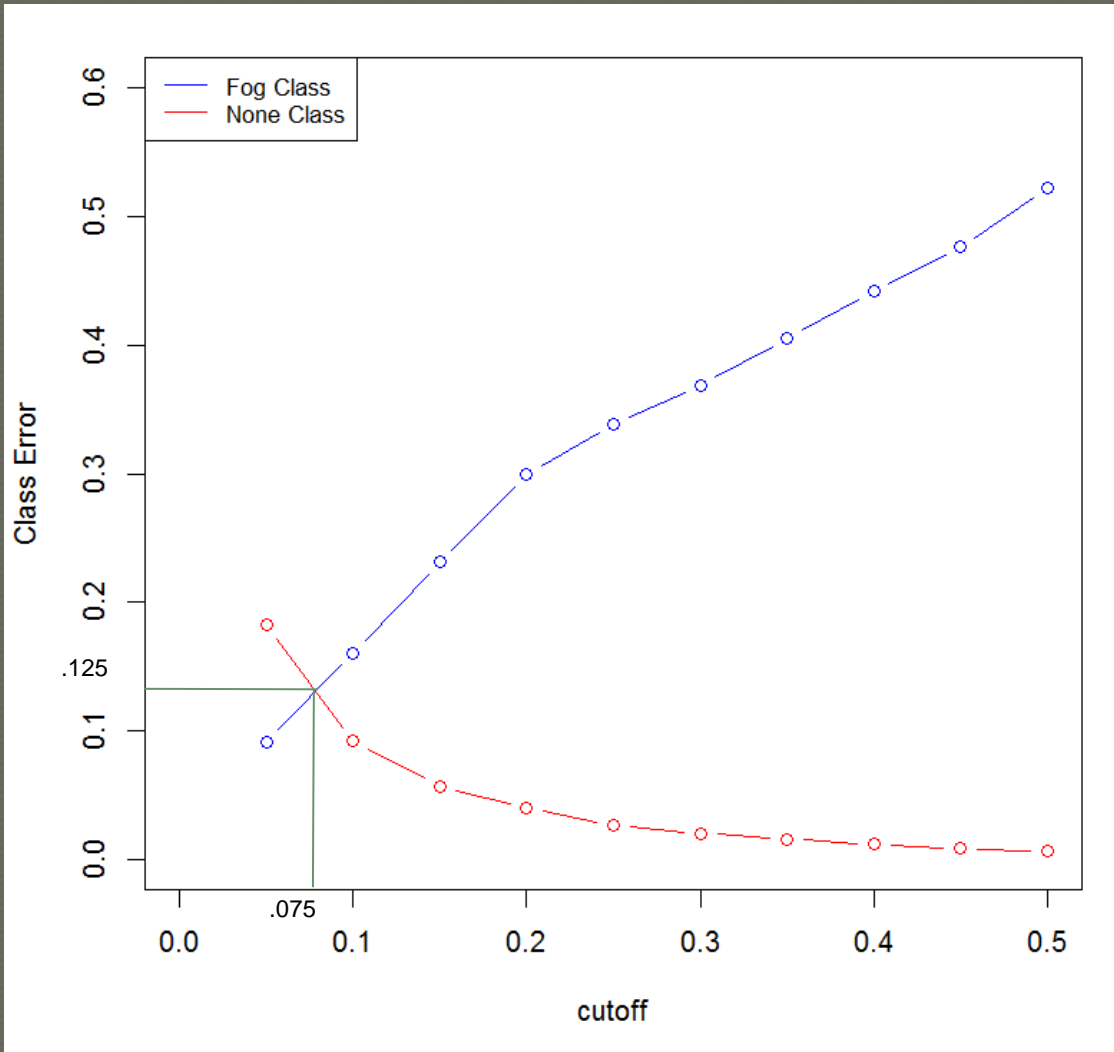
Predicted

Actual

	<b>Fog</b>	<b>None</b>	<b>Class Error</b>
<b>Fog</b>	352	387	.5237
<b>None</b>	63	10609	.0059



# Class Error vs. Cutoff





# Equal Class Error

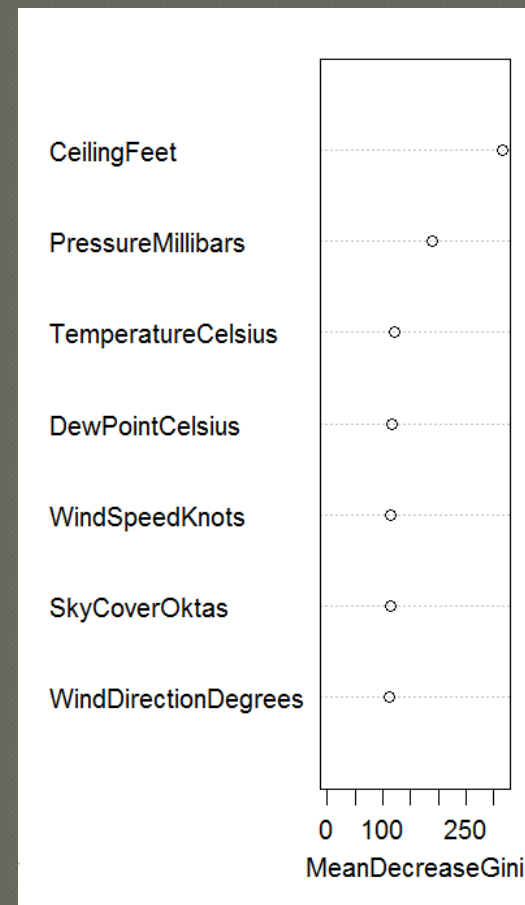
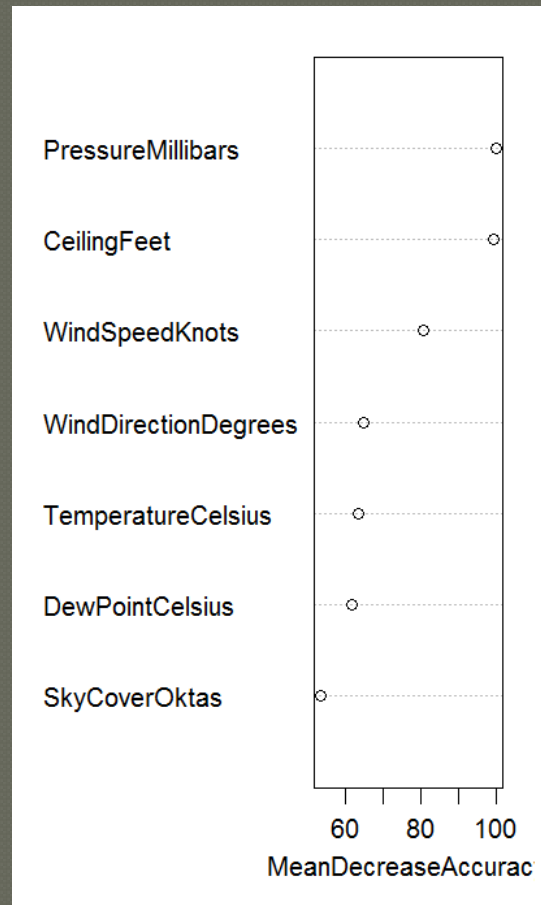
Predicted

Actual

	<b>Fog</b>	<b>None</b>	<b>Class Error</b>
<b>Fog</b>	644	95	.1286
<b>None</b>	1326	9346	.1242



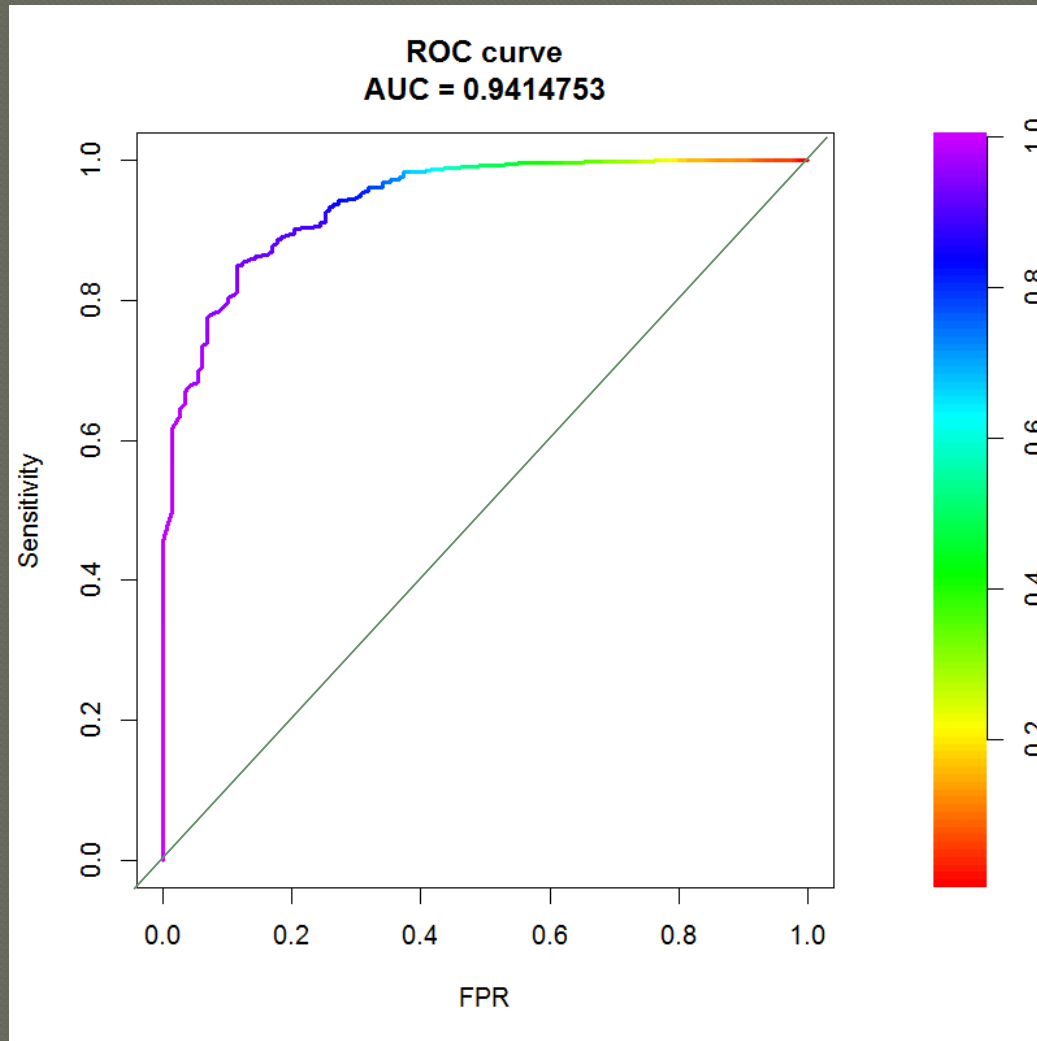
# Variable Importance





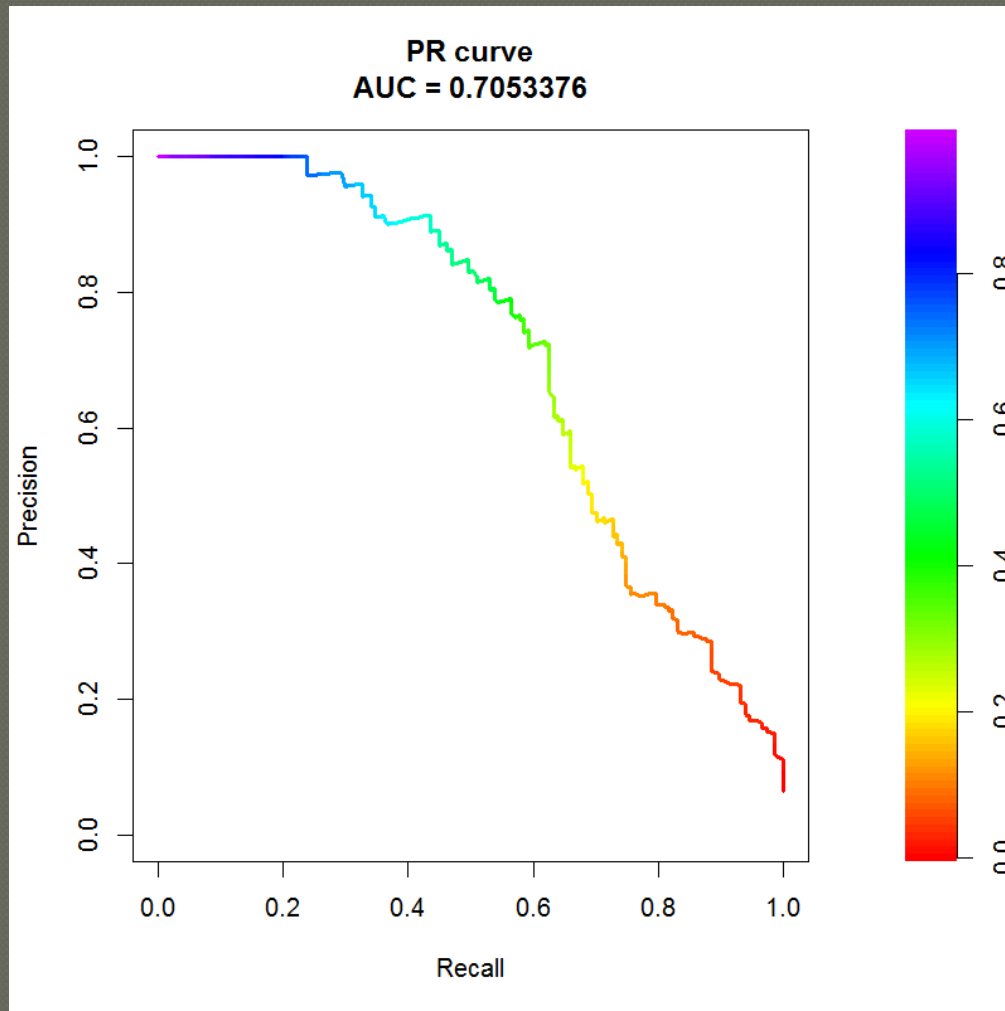


# ROC Curve





# Precision-Recall Curve





# 16/17 Random Forest

Predicted

Actual

	<b>Fog</b>	<b>None</b>	<b>Class Error</b>
<b>Fog</b>	43	24	.358
<b>None</b>	73	400	.154



# 16/17 Forecaster Average

Predicted

Actual

	Fog	None	Class Error
Fog	29.75	37.25	.556
None	116.75	356.25	.247



# Future Directions

---

- Add other predictor variables
- Optimize the Random Forest by tuning the cutoff based on a cost analysis
- Program the model into a forecasting application



# Acknowledgements and References

---

- ◉ SPAWAR Office of Polar Programs
  - ◉ Scientific Research Corporation
  - ◉ Mike Johnson, SOPP Chief Meteorologist
- 
- ◉ J. Davis, M. Goadrich, "The Relationship between Precision-Recall and ROC Curves", *Proc. Int'l Conf. Machine Learning*, pp. 233-240, 2006



# Questions?

If one decision tree forecasts incorrectly in the random forest and no one notices, does it verify?

