UNDERSTANDING HOW PROFESSIONAL PILOT TRAITS IMPACT PERFORMANCE

WHITE PAPER ON THE INSIGHTS OF MACHINE LEARNING AND PERFORMANCE PREDICTIONS

Author: Isaac Shareef, UAL 737 Flight Instructor
Date: May 4, 2021

Introduction:

Pilot shortages create very complex dynamics. As an airline with growth objectives, we will likely try to hire the best talent, and so will all of our competitors. If we are unlikely to hire the best and brightest all of the time, then it's best to prepare for a broader range of training outcomes. As airlines look to meet travel demand and growth expectations, how will Flight Training Departments handle the *known unknowns* in a cost sensitive environment? We know that newly hired pilots must have the minimum requirements for hiring, but we also *know* that it's *unknown* as to truly how capable they are for handling the rigors of flight training at a major airline. Are the current plans in place for training tomorrow's workforce data driven and cost effective?

Abstract:

Machine Learning(ML) can help our training department get a handle on the *known unknowns* by validating how various professional traits impact performance. What are the professional traits that have impacted pilot performance the most in our organization? The best indicator of performance from a cost perspective is a pilot that successfully completes training with no extra training events required. The professional traits of successful pilots are not random occurring, they are statistically significant. In an ideal scenario as the dynamics of hiring change, your training footprint will see no added costs from suboptimal trainee performance. If the ideal scenario is best hoped for, then utilizing ML becomes an excellent tool to prepare for anything less than ideal.

Background:

An in-house study was developed utilizing a ML model trained on the synthetic data of 17,000 pilots. The data set is comprised of 13 different professional attributes and the associated number of retrains per pilot. Below is a single row that shows how the dataframe was populated utilizing binary code.

Is Military	Is Civilian	Flew Fighter	Flew Cargo	Flew Corporate	Flew RJ	Arts Degree	Stem Degree	A-Hours	B-Hours	C-Hours	Prev Capt	Prev Rot-Wing
1	0	0	1	0	0	1	0	0	0	1	0	0

The dataframe was populated and iterated through to ensure a realistic data set was created that would likely resemble any air carriers pilot population. The target variable is the number of retrains. The retrain variables were synthesized into the dataframe and the entire dataset was used to train a ML model. After training the model we were able to learn the professional attributes that contributed most to retrains. The model enables single student retrain predictions, as well as batch predictions to determine how many retrains a random class could generate. The retrain predictions and the model gave incredible insights into performance and can have tremendous value with regards to cost and schedule planning. This study can be optimized by synthesizing real pilot data, and the model can be integrated into planning and decision making very seamlessly.

Conclusion:

The study conducted adds value by showing how a trained ML model can yield future performance predictions. After making predictions, best practices can be prescribed before training begins or during training to facilitate more successful and consistent training outcomes. The dynamics of tomorrow are very fluid, and ML can help us plan more effectively by being more aware, and more data driven.