Data Science
Computer Science

Capstone Report - Spring 2020

Dynamic Pricing for Commercial Aircraft

Yuxin Zhang

Supervised by
David Yu

Abstract

Efficient aircraft valuation and prediction are vital for Aircraft Appraisers, as well as other participants of the aviation industry. However, the monopolistic industry feature, pre-established theoretical valuation approaches, and lack of data transparency hindered the application of a quantitative modeling approach towards aircraft valuation. This project targets at solving those issues by pooling data from multiple sources such as reports from FAA, OEMs, Airlines, and Appraisers. Through a quantitative exploration of multiple models, including the Ridge Regression Model, the Random Forest Model, the XGBoost Model, this project eventually attains a Stacking Model, with Training_mse of 0.02145 and Testing_mse of 0.02159 (on Log(Base Value)/1 million dollar).

1 Introduction

The Aviation industry mainly consists of OEMs, Airlines, Leasing Companies, MROs, and Appraisers, with little data transparency between each other. Valuation for aircraft is an important issue for all participants in this industry, yet approaches and data sources differ due to diversity in interests and data availability. For example, Vasigh, Taleghani, and Jenkins suggest that DCF is a common approach for airlines to value aircraft from an operating cash flow perspective [1]. Gibson and Morrell also suggest that the NPV approach combined with real options analysis provides aircraft operators with flexibility in valuation [2].

Standing from the perspective of an Aircraft Appraiser, this project proposes to use a fundamental factor-based approach for aircraft valuation, which mainly considers the technical specifics (such as Cargo capacity, Range, Maximum takeoff weight (MTOW), seat numbers, etc) and operating specifics (depreciation, operating cost, fuel efficiency, maintenance cost, etc) [2][3][4][5].

The final delivery would be a "Base Value" model for commercial aircraft. This is proposed since definitions and factors concerned differ for different "value" under discussion. According to the definitions of "Value" proposed by ISTAT (International Society of Transport Aircraft Trading) [1], "Base Value" is an appraiser's option of the underlying economic value, with a reasonable balance of supply and demand. Besides, various quantitative modeling techniques would be examined on the dataset engineered from scratch. For the purpose of feature importance diagnosis, this project designs to examine the Ridge Regression Model, the Random Forest Model, the XGBoost Model, and apply a Stacking method to boost the general model performance.

2 Related Work

The aviation industry mainly consists of OEMs, MROs, Airlines, Leasing Companies, and Appraisers. OEMs design and manufacture aircraft and performance improvement packages (PIP) [6]; MROs maintain and repair aircraft; Airlines operate aircraft; Leasing Companies purchase and lease aircraft; Appraisers inspect and value aircraft. Although the value of an aircraft matters to all participants of this industry, the different business interaction with aircraft and market exposure determine the different approaches they would adopt in aircraft valuation.

Most modeling approaches are benchmarked from the aircraft operators' and investors' perspectives. Some popular proposed methods are net present value (NPV) approach, return on invested capital (ROIC) approach and the real option approach.

The net present value (NPV) approach takes on an operating cash flow perspective and is the most prevalently used valuation method for airlines and leasing companies [7]. Compared with the accounting-based approach, the cash-based approach reflects more of the real operating situation of the aircraft in use, directly linking to the real-time market and profit of operators. The general approach is to get a series of projected future cash flows generated by the aircraft and apply a discount factor to reach a present value [7]. However, this method has two flaws that limit its application in real practice – firstly, the geographic and economic features influence the adoption of this method. It's hard to predict operating cash in areas with many incentive investment plans such as tax credit, or banks being influential in investment decisions. Secondly, even with the rather flexible WACC-based NPV approach with Monte-Carlo Simulation, this valuation approach fails to capture the flexibility of operating leases and risks of owning and operating aircraft [2].

The return on invested capital (ROIC) approach, considering from the perspective of the long-term economic return, values the aircraft from a remarketing perspective. It includes the external market demand such as the number of aircraft in service, firm backlog, customer base, and position in the aviation cycle [5]. This approach incorporates the opportunities and risks of operating leasing and thus deals with the second issue presented by the NPV approach, as mentioned above. However, the cash flow estimation is still tricky since quantification for factors such as new market expansion, replacement and retirement decisions are still hard to make [8].

The real option approach is usually used by OEMs to assist with deciding which aircraft or performance improvement packages (PIP) to produce and how they should be priced. This is suggested since traditional cash-based valuation from an operating perspective fails to account for the flexibility offered to airline management to steer programs into profitable directions [6][9]. With a real option approach, it captures the opportunities for aircraft operators to expand new markets, develop new routes, form new fleet combinations, and increase competitive advantage with the purchase of new aircraft. This method better helps OEMs to position its product from the downward aircraft demand perspective. However, the possibility of each opportunity adopted in analysis in the real option approach is trivial to quantify and involves a lot of guesswork.

For the Appraising industry specifically, the quantitative valuation approach is rarely adopted. Available reference works focus on the valuation frameworks, measurement metrics, and researches on single factors' historical influence on aircraft valuation.

According to ISTAT (International Society of Transport Aircraft Trading), appraisers for most times deal with values, yet value definition differs according to their economic inclusion of factors. There are generally four different "Value" that appraisers refer to – the "Base Value", the "Market Value", the "Residual Value", and the "Distressed Value" (also refer to as "Forced Sale Value" or "Liquidation Value"). And the definitions are given as follows [10]:

- **"Base Value"** is the Appraiser's opinion of the underlying economic value of an aircraft in an open, unrestricted, stable market environment with a reasonable balance of supply and demand, and assumes full consideration of its "highest and best use". An aircraft's Base Value is founded in the historical trend of values and in the projection of value trends and presumes an arm's-length, cash transaction between willing, able and knowledgeable parties, acting prudently, with an absence of duress and with a reasonable period of time available for marketing. In most cases, the Base Value of an aircraft assumes its physical condition is average for an aircraft of its type and age, and its maintenance time status is at mid-life, mid-time.

- **"Market Value"** is the Appraiser's opinion of the most likely trading price that may be generated for an aircraft under the market circumstances that are perceived to exist at the time in question. Market Value assumes that the aircraft is valued for its highest, best use, that the parties to the hypothetical sale transaction are willing, able, prudent and knowledgeable, and under no unusual pressure for a prompt sale, and that the transaction would be negotiated in an open and unrestricted market on an arm's-length basis, for cash or equivalent consideration, and given an adequate amount of time for effective exposure to prospective buyers.

- **"Residual Value"** is the value of an aircraft, engine, or other items at a future date, often used in connection with the conclusion of a lease term.

- **"Distressed Value" (also refer to as "Forced Sale Value" or "Liquidation Value")** is the terms to describe the Appraiser's opinion of the price at which an aircraft (or other assets such as an engine or spare parts) could be sold in a cash transaction under abnormal conditions – typically an artificially limited marketing time period, the perception of the seller being under duress to sell, an auction, a liquidation, commercial restrictions, legal complications, or other such factors that materially reduce the bargaining leverage of the seller and give prospective buyers a significant advantage that can translate into heavily discounted actual trading prices. Depending on the nature of the assignment, the appraiser may be asked to qualify his opinion in terms of disposition within a specified time period, for example, 60 days, 90 days, or six months as the needs may be. Apart from the fact that the seller is uncommonly motivated, the parties to the transaction are otherwise assumed to be willing, able, prudent and knowledgeable, and negotiating at arm-length, normally under the market conditions that are perceived to exist at the time, not an idealized balanced market.

Vasigh's Book *Aircraft Finance: Strategies for Managing capital cost in a turbulent industry* sets framework for aircraft valuation and presents a collection of structured measurement prevalently used by the practitioner appraisers. In his book, three perspectives are suggested to be taken into consideration in terms of aircraft valuation, including the operational, financial, and technical perspectives. The operational dimension can be measured by factors such as average fuel burn (per block hour), average aircraft utilization (in block hours), and average stage length. The financial dimension can be measured by factors such as crew cost per block hour, depreciation per block hour, fuel cost, and maintenance cost per block hour. The technical dimension can be measured by factors such as average seats per aircraft, cargo capacity, range, and maximum takeoff weight (MTOW) [1].

Brewer and Cory Ryan, however, stress that appraisers focus on the measurement of the time factor as according to their paper *Techniques to Complete an Aircraft Appraisal*, the time factor has the largest influence on the "Base Value" of aircraft in two ways. Firstly, the structure of the depreciation of aging aircraft converges to the engine values at the end of its useful life, composing around 40% of the aircraft value. Secondly, the time information of a certain type of aircraft out of production would significantly decrease the value of the aircraft in the following years [11].

Besides the theoretical frameworks for aircraft valuation, some quantitative researches are done in exploring single factor's effect on aircraft value. The most researched factors are fuel price, new technology, and maintenance status.

The ICF report offers an industrial observation on fuel price's influence on the underlying economic value (the "Base Value") of aircraft. It shows, by presenting historical statistics, that fuel price comprises 14% to 31% of airline operating costs, thus being a key driver to purchase and retirement decisions for aircraft buyers. Furthermore, it shows a positive correlation between the fuel price and the number of new-engine adoption in aircraft transactions, which can be explained by the theory that airlines purchase more advanced aircraft to cut operating cost when fuel cost is high [3]. However, in another report published by ICF, *Why Aviation investing doesn't rely on fuel prices alone*, it shows that fuel price has a different effect on narrow-body and wide-body aircraft values. By stating that based on quantitative testing on historical data, it suggests that numerous aspects should be viewed interactively to determine the aircraft values [12].

In *The value of new-technology single-Aisles*, Steve Mason and James K.D. Morrison explain the aircraft operators' initiations for new-technology adoption. They suggest that cost containment, technology monopoly, regulation requirement, and new opportunities from expanding routes and utilizations are of major consideration when adopting aircraft of new design and engine type.

They specifically stress on the importance of engine types on the valuation of aircraft as it directly links to the range, the fuel efficiency of a certain aircraft [7].

In *The Relationship between an Aircraft's Value and its Maintenance Status*, Shannon Ackert states that maintenance status, by connecting with the operation of aircraft directly, influences both the "Base Value" and "Market Value" of aircraft. He also points out that the disparity in value in similarly aged aircraft lies within the difference in maintenance condition. Through studying the historical correlation of maintenance status with aircraft value, this article shows that depending on the aircraft type and age, the influence of maintenance status on aircraft value differs. Besides, it also defines the full-life status and half-life status assumptions that are prevalently used in aircraft modeling. The full-life status implies that each major maintenance event has just been fully restored or overhauled to zero time conditions. The half-life status, on the other hand, assumes that the airframe, engines, landing gear, and all major components are half-way between major overhauls and that any life-limited part (for example an engine disk) has used up half of its life [4].

3 Solution

1. Framework and Approach

   Different from some of the approaches discussed above, this project focuses on building aircraft valuation models from an appraisers' perspective, taking dynamics in the whole aviation industry into consideration. Before exploring the framework of model establishment, it's better to take a stand on what "Value" is being modeled in this project, as there are four most commonly referred-to "Value" for appraisers, as discussed above [10]. According to definitions provided by ISTAT (International Society of Transport Aircraft Trading), "Base Value" is an appraiser's opinion of the underlying economic value of an aircraft in an open, unrestricted, stable market environment with a reasonable balance of supply and demand. Besides assuming the "highest and best use" of aircraft, it also makes the assumption that each aircraft is in the average condition of its type and age, and has half-life maintenance status, which is an extremely important and reasonable assumption for benchmarking the valuation model compared to the full-life status assumption as few situations would an aircraft not get maintained until fully depreciated [4].

   Due to the lack of published research on aircraft valuation modeling and the monopolistic industry feature, this project approaches the modeling by firstly structuring the theoretical framework. Two general perspectives – the technical, and the operational, are taken into consideration. The technical perspective involves specifications of aircrafts such as seat numbers, range, fuel capacity, and maximum takeoff weight (MTOW), which generally come from the OEMs records. The operational perspective includes features such as average fuel burn, and utilization of aircraft, which mainly come from aircraft operators.

   Upon structuring the valuation framework from two perspectives, this project takes a substantial effort in feature engineering, with an inclusion of many historical industrial observed relevant features, as discussed above. For example, fuel cost, which composed 14% to 31% of airline operating cost, is an important factor in aircraft valuation [3]. And it is a factor both influenced by the market condition and the technical specifics of an aircraft. However, historical evidence shows a strong signal that fuel price has a different effect on the orders for narrow-body and wide-body aircraft, which indirectly affects "values" accordingly.

This, in theory, has to do with the different customer base, route fragmentation and market sentiment of narrow-body and wide-body aircraft demands. It suggests that the inclusion of body type into valuation modeling is necessary to quantify the different effect of fuel price on aircraft valuation [12]. Also, another observation shows consistent convergence of aircraft value to its engine value towards the end of an aircraft's useful life, which suggests that engine type is an important factor in time-involved valuation [11]. Besides, strong evidence shows that maintenance cost is subjective to market forces including the inflation rate, which over the long-term has the potential to make an aircraft uneconomical to operate. That's why this project makes the assumption of baseline inflation at 3% , upon which offers flexibility to make a change.

In summary, this project innovatively adopts a method that supports quantitative valuation modeling while utilizing a theoretical framework adopted by practioner appaisers in feature engineering. Besides, this project establishes solid assumptions on valuation modeling regarding factors such as inflation rate and half-life maintenance status for flexible adjustment.

2. Data Engineering and Processing

   This project engineers data from scratch. There are three parts of the data to be engineered according to the theoretical framework. The first is the operating factors such as fuel cost, maintenance cost, and load factors. They are engineered from the annual financial report from Airline companies and FAA, and averaged-out on annual basis for each type of aircraft. The second is the technical factors such as Cargo capacity, Range, Maximum takeoff weight (MTOW), and seat numbers. They are engineered from OEM companies' annual reports. The third is the historical "Base Value" recorded by Appraising Companies. Four Appraising Companies (IBA, AVITAS, BACK/CV, MBA) are taken as the raw data sources because they take slightly different approaches and record data at different time point of the year. After each part of the data taken, a preliminary cleaning is done on "Plane Type" and "Sub-Type" to ensure the accurate match when joining the three datasets. This has to be done because different companies take different standards in the recorded naming for the same aircraft, especially for the "Sub-Type" recordings. After aligning the "Plane Type" and "Sub-Type", a merge of three datasets into one complete datasheet is done for following steps.

   Then a general data cleaning is done from two perspectives. Firstly, outliers for "Base Value" are detected and cleaned for a certain type of aircraft delivered in the same year. Secondly, an economic logic check is done for the "Base Value" of aircraft. This is to ensure that the value for one type of aircraft always goes down with time and that earlier-delivered aircraft worth less than the later-delivered ones.

   After the data cleaning, three new data columns are generated from the existing ones to provide important extra information to the dataset for better modeling. A categorical variable "Passenger/Freighter" is generated based on "Plane Type" to indicate the useful purpose for the aircraft. Another categorical variable "Body type" is also generated based on "Plane Type" to indicate the physical size and capacity of the aircraft. Besides, a numerical variable "Age" is calculated by the difference between the time when the value is measured and the time of delivery to measure the depreciating effect of time.

   With all columns prepared, a feature selection on numerical factors is done in two steps to ensure that each column to be included in the model possesses enough information value. Firstly, a correlation heap map is generated to check the correlation between the independent variables and the dependent variable ("Base Value") [see Figure 1]. Theoretically, independent variables should have little correlation between each other so that parameter precision after model training won't suffer from inflated variance caused by the multicollinearity issue. Also, the independent variables should have a high correlation with the dependent variable to be informative in model construction. According to such criteria, several variables are dropped, resulting in only eight numerical variables. Secondly, from the model contribution perspective, the Recursive Feature Elimination (REF) method from the Python Sklearn package is used to select the best feature combinations. This method takes a recursive way to remove the weakest features. And by applying it, all eight numerical features show statistically significant contribution to the model.

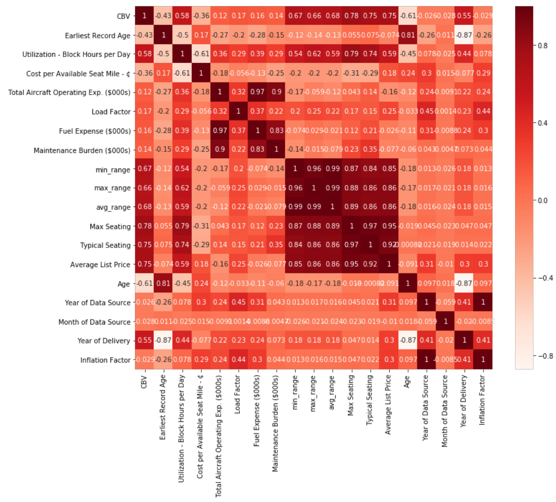

Figure 1: Correlation Matrix for Numerical factors

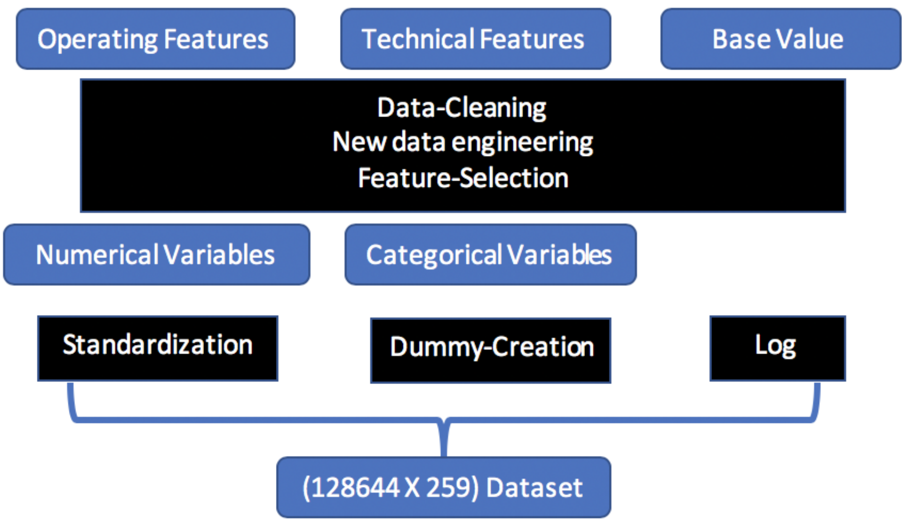

Figure 2: Data Processing Logic Demonstration

Proceeding from feature selection, dummy creation is conducted for all categorical independent variables including "Aircraft Type", "Sub-Type", "Body Type" and "Passenger/Freighter". Also, for the eight numerical independent variables, standardization is performed to regulate their distribution for better modeling precision purposes. In addition to the independent variables, the dependent variable "Base value" is done a distribution adjustment by taking a log form [see Figure 2].

In the final step of data processing, a random 80-20 Train-Test Split is performed to attain the same training and testing set for the all models being explored in the following steps. In this way, a benchmark is set for the model performances to be comparable.

3. Model Training and Parameter Tuning

After the Data Engineering and Processing, the dataset has shape of (128644, 259), with training set of shape (102915, 259) and testing set of shape (25729, 259).

The aircraft valuation models are generally considered as a prediction model. And this project considers training models with good interpretability, as it is expected to fit with the underlying theoretical frameworks. Attempting to capture the most features out of the dataset, this project targets to utilize the ensemble methodology, including the bagging, boosting, and stacking. That's why in addition to the most straightforward Regression Model, it also applies Random Forest Model, which is the bagging algorithm, and the XGBoost Model, which is a boosting algorithm. Upon all three models trained separately, it also targets at stacking them together for a more comprehensive Stacking Model. This, in theory, would produce statistically and computationally better performance on both the training and testing dataset.

a) Ridge Regression Model

Ridge Regression Model is one of the Regression models used for analyzing feature importance. Different from the Ordinary Least Square Regression, which doesn't differentiate "important" from "less-important" predictors in a model, Ridge Regression, by posing a "regularization" on the optimizer regarding the added predictors, avoids the overfitting problem [see Figure 3]. Another advantage of the Ridge Regression is that it is less restricted in the assumption as it doesn't require completely unbiased estimators. Besides, the Ridge Regression is flexible in application as it performs well even when multicollinearity issue in estimators is present.

$$\mathcal{J}_{\text{SRM}}(f) = \underbrace{\frac{1}{M}\sum_{i=1}^{M}\ell(y^{(i)}, f(\mathbf{x}^{(i)}))}_{\text{risk/data fit } R_{\mathcal{D}}(f)} + \underbrace{\lambda\,\Omega(f)}_{\text{regularization}}$$

Figure 3: Logic Demonstration for Ridge Regression model

Only one parameter Alpha, which controls the strength of the "regularization"(or "penalty") term in Ridge Regression, is tuned using GridSearchCV method from Python Sklearn package. GridSearchCV generally exams a list of Alpha values and return the one with the best training and validation result. Finally, the resulting best parameter Alpha (which is 0.1 in the case) is viewed as the tuned parameter, which is ready for applying in model training. Using MSE to measure the model performance, the Ridge Regression attains a Training_MSE of 0.08708 and a Testing_MSE of 0.08536.

Analyzing the trained Ridge Regression Model with tuned best parameters, the Dummy Variable "Plane Type" ranks the highest in feature importance [see Figure 4].

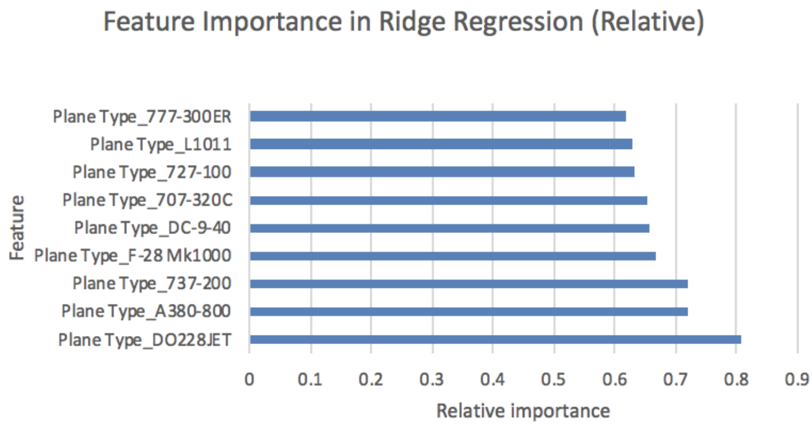

Figure 4: Importance Factor Ranking for Ridge Regression Model

b) Random Forest Model

Random Forest is an ensemble technique that uses multiple decision trees and a bagging technique, which involves training each decision tree on a different data sample where sampling is done with replacement. One of the most salient advantages of the bagging method is that it reduces the variance of the model. And the Random Forest, by sampling a reduced number of features in each decision tree, encourages de-correlation between different individual trees as it discourages some particularly strong predictors for the target variable to be included in every individual tree [see Figure 5].

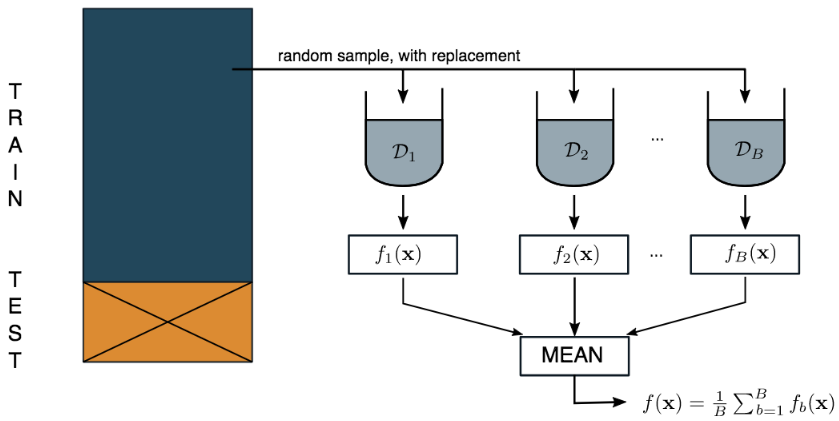

Figure 5: Logic Demonstration for Bagging Ensemble method

A bunch of hyper-parameters are tuned before training. For example, parameters such as the number of trees in the forest, the maximum depth of the tree, the minimum number of samples on leaf node, and the number of features for best split. A Python Sklearn method called RandomizedSearchCV is applied in searching for the best parameter combination that produces the minimum MSE on the validation set. In contrast with the GridSearchCV where all parameter values are tried out, the RandomizedSearchCV randomly draws parameters from a given distribution.

Then 100 times random parameter drawings are performed, each time training the model on a 3-fold basis. Finally, the resulting best parameter combination is viewed as the tuned parameters, which are ready for applying in model training. Using MSE to measure the model performance, the Random Forest Model attains a Training_MSE of 0.0312 and a Testing_MSE of 0.03519.

Analyzing the trained Random Forest Model with tuned best parameters, the Dummy Variable "Body Type", and "Plane Type" rank the highest in feature importance [see Figure 6].

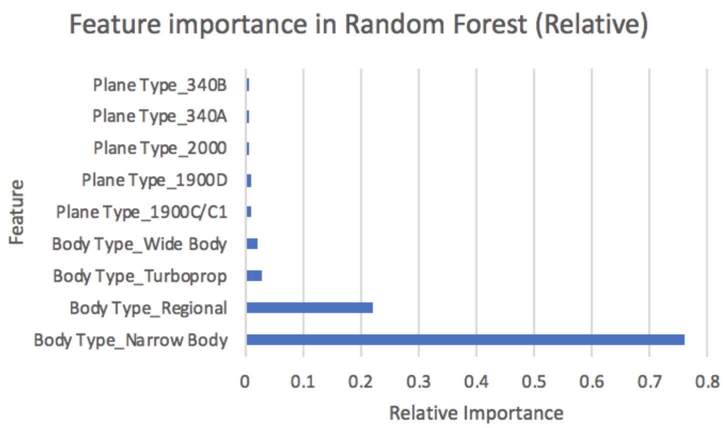

Figure 6: Importance Factor Ranking for Random Forest Model

c) XGBoost Model

XGBoost is one of the boosting ensemble method. A boosting algorithm improves the model performance by training a set of weak learners of high bias and low variance, which would eventually obtain one strong meta-learner. Yet variance would increase as the performance of the model improves. The XGBoost Model, possessing the general features of a boosting method, has its own advantages. Firstly, XGBoost has in-built L1 (Lasso Regression) and L2 (Ridge Regression) regularization which prevents the model from overfitting. Secondly, XGBoost utilizes the power of parallel processing and thus is much faster as it uses multiple CPU cores to execute the model. Thirdly, XGBoost has an in-built capability to handle missing values. In fact, when XGBoost encounters a missing value at a node, it tries both the left and right hand split and learns the way leading to higher loss for each node, which would be helpful in handling some missing values regarding operating features in the dataset of this project [see Figure 7].

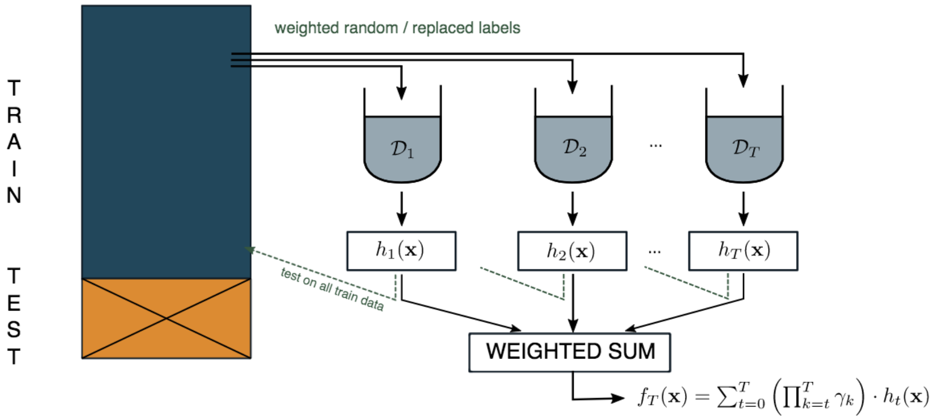

Figure 7: Logic Demonstration for Boosting Ensemble method

A bunch of hyper-parameters are tuned before training. For example, parameters such as the type of booster, the learning rate, the maximum depth of a tree (if applying tree booster), the subsampling parameters, and regularization terms. A Python Sklearn method called RandomizedSearchCV is also applied in searching for the best parameter combination that produces the minimum MSE on the validation set.

Then 100 times random parameter drawings were performed, each time training the model on a 3-fold basis. Finally, the resulting best parameter combination is viewed as the tuned parameters, ready for applying in model training. Using MSE to measure the model performance, the XGBoost Model attains a Training_MSE of 0.03661 and a Testing_MSE of 0.03722.

Analyzing the trained XGBoost with tuned best parameters, the Numerical Variable "avg_range_SD", and "Year of Delivery_SD" rank the highest in feature importance [see Figure 8].

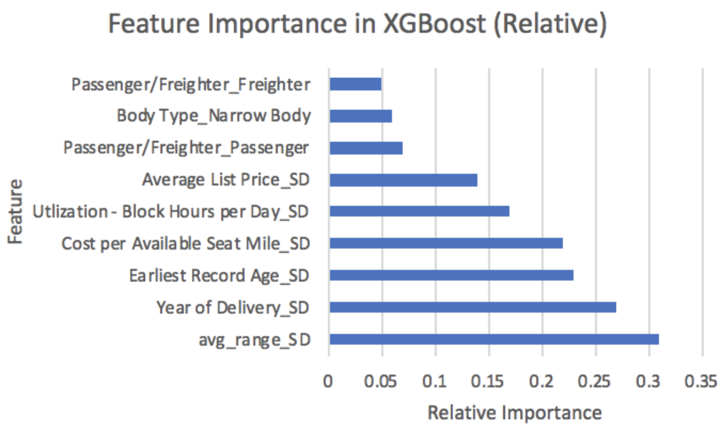

Figure 8: Importance Factor Ranking for Xgboost Model

d) Stacking Model

A stacking model is a Meta model trained upon a bunch of trained models. By doing so, it aims to capture the best part of each model. In this way, the stacking method promises to produce a better performance than each individual models [see Figure 9].

$$f_1(x) = RidgeRegressionModel$$

$$f_2(x) = RandomForestModel$$

$$f_3(x) = XGBoostModel$$

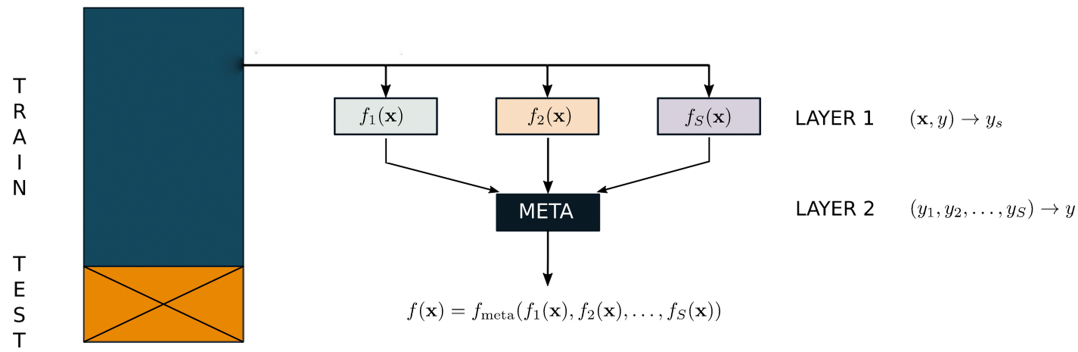

Figure 9: Logic Demonstration for Stacking Ensemble method

By using StackingRegressor method from Python Sklearn package, a stacking model is built upon the previous three trained models (the Ridge Regression model, the Random Forest model, and the XGBoost model). All trained models are inputted with their tuned parameters and a GradientBoostingRegressor is applied in the Meta Model training.

The Stacking Model produces better MSE performance compared to the that of individual models. By tuning the parameters such as the learning rate and the maximum tree depth, the Meta-model is eventually trained with the tuned parameters, producing Training_MSE of 0.02145 and Testing_MSE of 0.02159.

4 Results and Discussion

1. Result assessment

   a) Model performance and comparison

This project uses MSE on both training set and testing set to measure the performance of the model. The Training set is used in parameter-tuning and model-training, with the testing set only utilized after the model is developed. With both the Training_MSE and Testing_MSE, it's easy to diagnose over-fitting and not fitting for the model constructed. And it also provides a measurement that's comparable between different models, with lower MSE indicating better model performance. The Training_MSE and Testing_MSE result for all four models involved in the previous discussion are as follows [see table 1].

For the three individual models, the Random Forest and XGBoost has much better performance than the Ridge Regression. And the Stacking Model, as suggested by theory, does indeed do better than any of the three individual models. And from its resulting similar Training_MSE and Testing_MSE, the Stacking Model is considered to have no overfitting issue, thus is the best model fitting the dataset.

	Ridge Regression	Random Forest	XGBoost	Stacking Model
Training_MSE	0.0871	0.0312	0.0366	0.0215
Testing_MSE	0.0854	0.0352	0.0372	0.0216

Table 1: Model Performance Table

Plotting the Training Models' Performance on the entire dataset, it's clearer to see how the Random Forest Model, the XGBoost Model, and the Stacking Model perform comparatively better by contracting the predicted value towards the measured value. The fitness (measured by R^2) also tell the same story [see Figure 10].

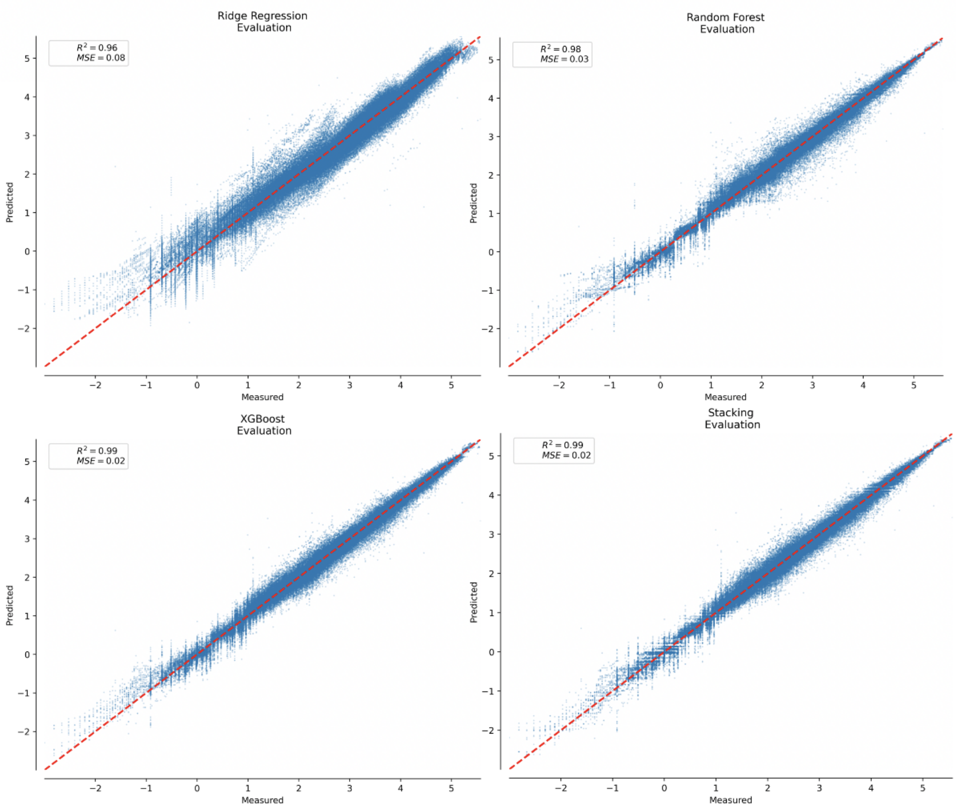

Figure 10: Model Performance Evaluation on Full Dataset

b) Explanation for the factor importance

It's expected that top contributors from different models show similar patterns that fit the theoretical framework. Yet as shown in the previous important factor rankings, the top-ranking important factors in XGBoost Model and Random Forest model show differences. So it's neccesary to check the correlation between the top ranking factors' distribution, which is between the "Body Type", a categorical feature, and "avg_range", a numerical feature.

Using one-way analysis in scipy.stats between variable "Body Type" and variable "avg_range", it attains the F_onewayResult with statistic of 7280, and p_value of 0.0, showing an overall significant effect of "Body Type" on "avg_range". In this way, the difference in feature importance ranking can be explained by the correlation instead of difference in model-recognized pattern. This fits what the theory tells as well, as the "Body Type" defines generally the technical size and operational capacity of a plane, thus is highly correlated with the range of the aircraft.

2. Challenges of the project

a) "Base Value" Extraction and Processing

The raw data sources are reports from four Appraiser Companies (IBA, AVITAS, BACK/CV, MBA). Four issues exist in the raw data extraction and processing from different sources.

Firstly, "Plane Type"' and "Sub-Type" are not recorded in the same format for different sources. For example, for the "Sub-Type" for "Plane Type" of "727-200A", "JT8D-9/-9A" is sometimes recorded as "FedE Stage 3 Kit", or "Nordam Stage 3". This is dealt with by referring to a full list of all commonly used "Plane Type" and "Sub-Type" combination after attaining data from all sources.

Secondly, for "Base Value" recordings for certain type of Planes, the "Year of Delivery" is recorded as "Typical" for simplicity reasons. This issue is dealt with after attaining all data from all sources. All Planes with "Year of Delivery" recorded as "Typical" are duplicated, with "Year of Delivery" replaced with a list of its typical delivery year from other value-measured year or from other sources.

Thirdly, different sources have different months of data recording. For example, for "AVITAS" data source, it records values twice a year in March and September. Plotting values for one certain aircraft (decided by "Plane Type" and "Sub-Type") by year, annual variance is too large and part of time factor's influence on value is lost. For example, the graph below shows the annual "Base Value" distribution for aircraft "717-200" with no "Sub-Type" delivered in 1999 [see Figure 11]. Although interval for value recording is not constant, this project decides to include month factor to further capture the time influence for value. Thus a much smooth trend can be perceived for aircraft "717-200" with no "Sub-Type" delivered in 1999 [see Figure 12].

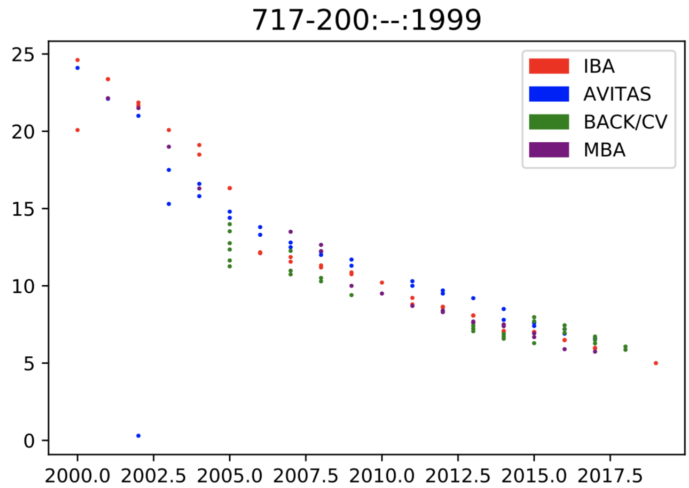

Figure 11: Base Value Annual distribution for 717-200 delivered in 1999

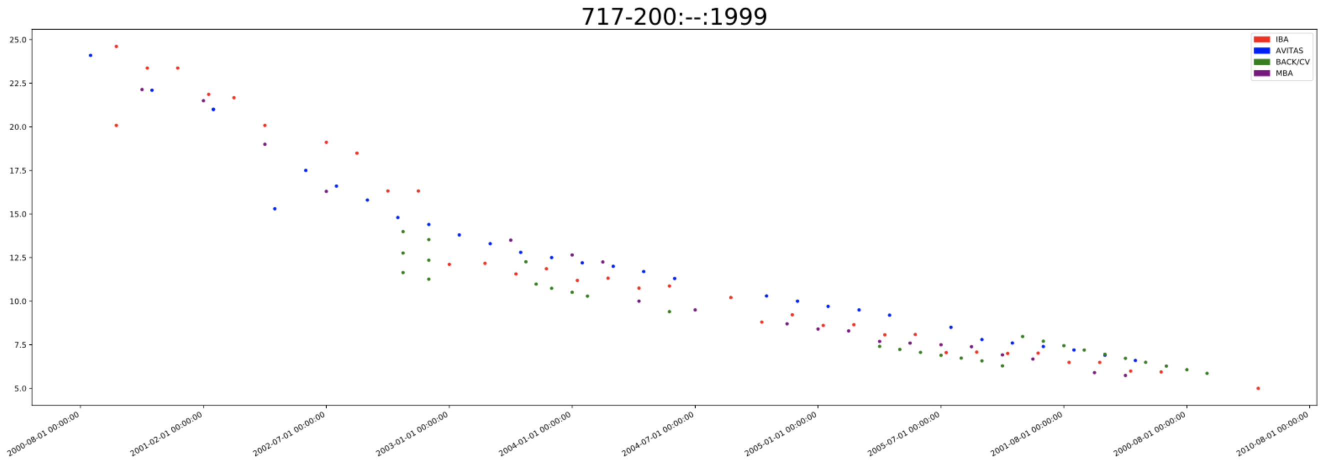

Figure 12: Base Value Year-Month distribution for 717-200 delivered in 1999

Fourthly, as raw PDF reports are converted using a software called FineReader, converting issues, though rarely occur, do exist. Refering to the value distribution graph of aircraft "717-200" with no "Sub-Type" delivered in 1999 shows, a value outlier shows in the year 2002 from source AVITAS [see Figure 11]. The complexity issue with dealing with an outlier of aircraft is that a huge drop in value does exist between monthly intervals due to economic reasons. This is solved by first identifying the outliers from the statistics perspective. Then manual reference is made with the original PDF source to verify the real outliers. The above example is an outlier and it is caused by the FineReader misplacing the decimal point when doing the file format conversion task.

b) Choice of model

i. Why not LSTM

LSTM is primarily considered as it is a flexible RNN algorithm designed for prediction or classification when things have time series dependency features. But its application has some assumptions that are not satisfied in the dataset. If to use LSTM, this project sets the assumption that:

- An average of data points from different sources measured at the same time point is taken
- The time interval between the different points is regraded the same

In this way can this project use the past certain amount of data points to predict the future ones. When trying this methodology, this project takes the past 300 data points to predict the next 300 ones. And the detailed steps are as follows:

In step one, dataset (before the Train-Test Split in the data processing step) is ranked by "date" , the time of value measurement. The Train-Test Split is done by taking the first 80% of ranked dataset as training data. In step two, X_train and Y_train dataset are built, with X_train includes the past 300 data points, and Y_train includes the next 300 data points. X_train and Y_train are built upon the training set taken in Step one. Both X_train and Y_train are arrays of 3 dimensions. In step three, a random shuffle is performed on the first dimension of X_train and Y_train, taking the same order. In step four, 10% of both X_train, Y_train are taken as validation set (X_val and Y_val) in model training. In step five, Many-to-Many Model in LSTM are applied. By using "MSE" for loss measurement and "adam" as an optimizer, an LSTM model is trained, with validation loss measured on the validation set. Eventually, the Model reaches a Traing_mse of 0.1159 and a Testing_mse = 0.1167.

This is not a good result but is expected since the model is forced to view time-interval between different data points to be the same, yet they are not. In fact, the time-interval for all data has the following distribution [see Figure 13]. Also, by taking average value from different sources at the same time point, partial information is lost. Due to the unsatisfying dataset condition for LSTM application, the LSTM algorithm is not taken into consideration by this project.

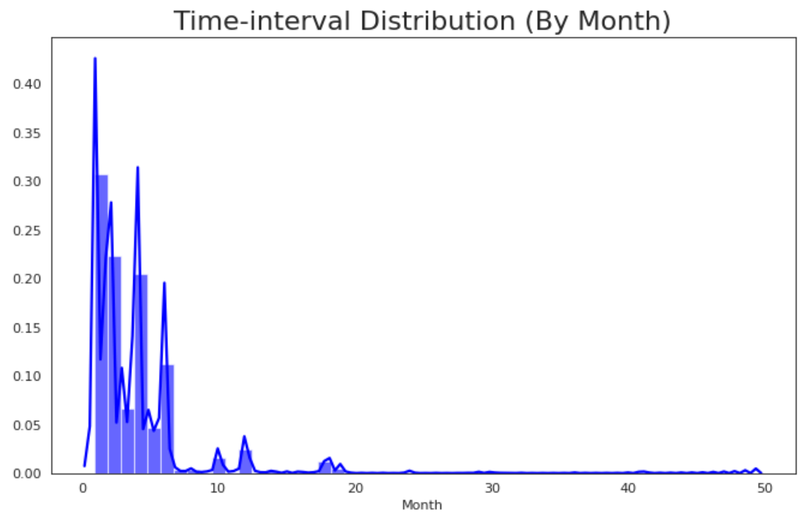

Figure 13: Time-interval distribution (By Month)

ii. Interpretable models and Ensemble method

As an industry-oriented project supported by the practitioner appraiser valuation framework, it is expected to provide explanatory power to variable importance in the model. That's why to better capture the feature in the dataset and boost performance, this project targets as using a Stacking method as a general approach. After parameter-tuning and training for the Ridge Regression Model, the Random Forest Model, and the XGBoost Model, this project stacks three models together for a Meta Model. In this way, it can produce better results overall due to reasons from three perspectives. From the statistical perspective, averaging models reduces the risk of high variance problems resulted from training data being too small compared to the hypothesis space. From the computational perspective, learning algorithms based on gradient descent in non-convex spaces or other greedy algorithms can get stuck in local optima, even when enough data is provided. Ensembles start from different points in slightly different landscapes and may when averaged better approximate the true underlying function. Besides, from the representational perspective, it may be possible to expand the hypothesis space by combining more models [see Figure 14].

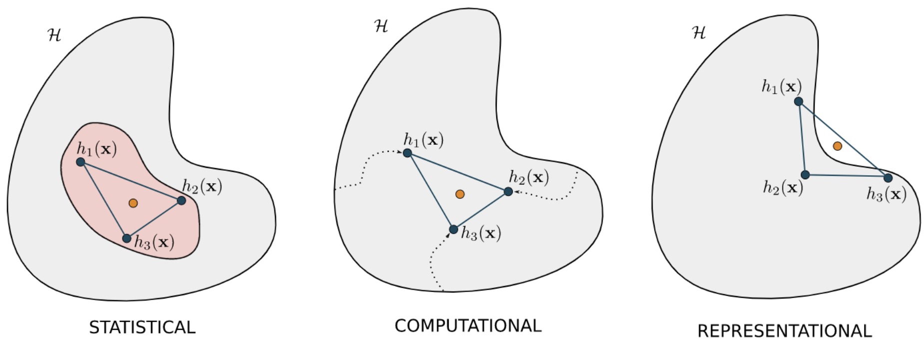

Figure 14: Demonstration of Benefit for Stacking Method

c) Deal with two Time-Dimension features

For aircraft of each "Plane Type" and "Sub-Type" combination, its value is recorded according to two dimensions regarding time – "Year of Delivery" and "Time of Value Measured" (Jointly measured by factors "Year of Value Measured" and "Month of Value Measured"). This is problematic since this project does not intend to build a separate model for each aircraft delivered at different years.

To solve this problem, a column "Age" is calculated as a difference between "Time of Value Measured" and "Year of Delivery". In this way, the dataset has only the "Year of Delivery" as a time dimension and has converted another time dimension to a numerical variable "Age". Theoretical this makes sense as all aircraft should have a general trend of value depreciation when "Age" goes up. A depreciation chart for all value data points and a general fitted line is graphed to verify this [see Figure 15]. With the y-axis being the % of its value at the delivered year, and the x-axis being "Age". The general depreciation effect of "Age" for value of any type of aircraft is verified by the existing dataset as well .

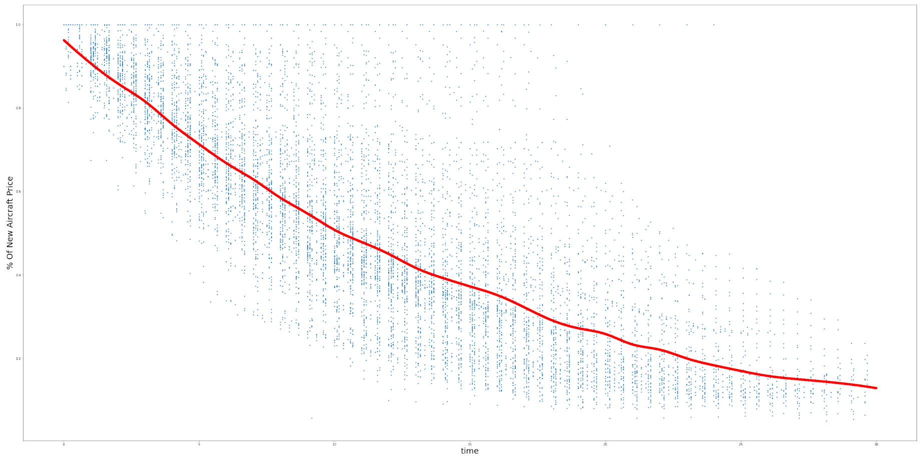

Figure 15: Age depreciation for all aircraft

5 Conclusion

1. Contributions

   - By taking the effort in acquiring and engineering data from multiple sources, this project succeeds in exploring a quantitative approach in aircraft valuation

   - By spending large amount of time doing data cleaning, this project achieves the maximal usage of raw data while making sure the data is intact and unbiased

   - By utilizing the fundamental factor-based framework adopted by practitioner appraisers and applying interpretable models on the dataset, this project succeeds in bringing in interpretability to the quantitative model for theory reference and verification

   - By calculating "Age" for aircraft and replacing one time-dimension measurement with it, this project successfully handles issues with two time-dimensions modeling

   - By adopting Ensemble methods such as Bagging, Boosting, and Stacking to the model construction process, this project makes great effort to fully capture the features in the dataset, and thus boosting the model performance

2. Direction for improvement

   - More variables theoretically influencing the "Base Value" for aircraft can be added to boost model performance. For example, in theory, the number of available engines in a certain aircraft can influence its value due to the added value in the given customization right.

   - Inter-time value differences can be a direction for exploring other than modeling value itself

   - Quantitative models for other "Value", especially the "Market Value", can be a wild field for exploring, as economic and market data is both comprehensive and easy-accessible

References

[1] B. Vasigh, R. Taleghani, and D. Jenkins, *Aircraft Finance:Strategies for Managing Capital Costs in a Turbulent Industry.* J. Ross Publishing, 2011.

[2] W. Gibson and P. Morrell, "Theory and practice in aircraft nancial evaluation," *Journal of Air Transport Management*, vol. 10, pp. 427–433, 2004.

[3] S. Rubin and L. Ranson, "Icf: Exploring fuel prices and the other factors behind aircraft aircraft values," 2019.

[4] S. Ackert and W. Gibson, "The relationship between an aircraft's value and its maintenance status," 2013.

[5] S. Ackert, "Basics of aircraft market analysis," *AIRCRAFT MONITOR*, 2013.

[6] W. Gibson and P. Morrell, "Airline finance and aircraft financial evaluation: evidence from the field."

[7] S. Mason and J. K. Morrison, "The value of new-technology single-aisles."

[8] J. Gorjidooz and B. Vasigh, "Aircraft valuation in dynamic air transport industry."

[9] C. Y. Justin and D. N. Mavris, "Valuation of real options with flexible early exercise in a competitive environment: The case of performance improvement packages."

[10] P. Morrell and W. Gibson, "Airline finance and aircraft financial evaluation: evidence from the field," *ATRS World Conference*, 2005.

[11] Brewer and C. Ryan, "Techniques to complete an aircraft appraisal."

[12] ICF, "Why aviation investing doesn't rely on fuel prices alone."