Submitted to *INFORMS Journal on Applied Analytics*
manuscript (Please, provide the manuscript number!)

Galactic Air* improves ancillary revenues with dynamic personalized pricing

Arinbjörn Kolbeinsson
Imperial College London, UK, ak711@imperial.ac.uk

Naman Shukla
Deepair Solutions, naman@deepair.io

Akhil Gupta
Industrial and Enterprise Systems Engineering, University of Illinois at Urbana-Champaign, akhilg3@illinois.edu

Lavanya Marla
Industrial and Enterprise Systems Engineering, University of Illinois at Urbana-Champaign, lavanyam@illinois.edu

Kartik Yellepeddi
Deepair Solutions, kartik@deepair.io

Ancillaries are a rapidly growing source of revenue for airlines, yet their prices are currently statically determined using rules-of-thumb and are matched only to the average customer, or to customer groups. Offering ancillaries at dynamic and personalized prices based on flight characteristics and customer needs could greatly improve airline revenue and customer satisfaction. Through a startup (Deepair) that builds and deploys novel machine learning techniques to introduce such dynamically priced ancillaries to airlines, we partnered with a major European airline, Galactic Air, to build models and algorithms for improved pricing. These algorithms recommend dynamic personalized ancillary prices for a stream of features (called *context*) relating to each shopping session. Our recommended prices are restricted to be lower than the human-curated prices for each customer group. We designed and compared multiple machine learning models and deployed the best performing ones live on the airline's booking system in an online A/B testing framework. Over a six-month live implementation period, our dynamic pricing system increased ancillary revenue per offer by 25% and conversion rate by 15% compared to the industry standard of human-curated rule-based prices.

Key words: airline ancillary pricing, dynamic pricing, customized pricing, deep learning
History:

Airline operations is a capital intensive and low margin business. Southwest Airlines pioneered the Low Cost Carrier (LCC) business model in the 1970s with the sole objective to offer cheaper airfares to customers. Low cost airlines are able to offer competitive prices by adopting one or more of these business model levers - (1) they keep their operating and fixed costs low by operating from underused (less expensive) airports (2) they keep their

* pseudonym

distribution model simple and focus purely on direct sales through their website or call center. (3) they un-bundle their services in such a way that the base product only provides a bare minimum set of services, including the right-to-fly. They offer all additional services for an additional fee as optional add-ons. Un-bundling can be an effective strategy for both airlines and travelers. It helps airlines monetize value-added services and gives the travelers the option to only pay for things that they want to use in that particular trip.

Most full-service airlines have now evolved into a hybrid business model. They achieve this through product differentiation, also known as branded fares. Pioneered by Air Canada (Taneja 2016), branded fares are product bundles mainly differing in value-added ancillaries services which are packaged with a "right-to-fly" service. Ancillary revenues have been growing steadily since their introduction in mid-1990s. In 2018, airlines made $92.9 billion in ancillary revenue and it has been growing at the rate of 18% year-over-year (IdeaWorks 2015). While ancillaries represent only 11% of the total industry revenue, for some airlines such as RyanAir (Europe) and Spirit Airways (US) they represent over 22% of the airline's revenue. As competition increases, airlines will be forced to drop the price of their base product, and rely on effective upsell and personalization. This, in turn, will make ancillaries more important for the bottom line of the airline.

Personalized dynamic pricing is a widely used pricing strategy in many industries. The underlying motivation is straightforward: every customer has a different willingness to pay, an optimal pricing strategy should be able to match the price to each customer's maximum willingness to pay. There are many factors (also known as *context*) such as the competitor's price, product attributes and external market factors that could influence this willingness. Instead of offering the same price to everyone, the supplier uses algorithms or heuristics that would dynamically change the price based on the context. While the airline industry pioneered the practice of dynamic pricing in the 1980s, the most recent innovations have been in the ride-sharing industry, spearheaded by Uber and popularly known as Surge Pricing.

Dynamic Pricing in the Airline Industry

Today, most airlines use dynamic pricing that optimizes the price of the base product by maximizing the expected revenue (Belobaba et al. 2015, van Ryzin and Talluri 2005). The system's goal is to serve the right base product price, to the right customer at the right time. Dynamic pricing of the base product typically takes into account the context of the available supply, demand forecast, and willingness to pay of the demand that is yet to come. Some research studies the effects of considering future expected ancillary sales into the pricing of the base product, but very little research focuses on the dynamic pricing of the ancillaries.

Over the past three decades, while airline pricing and revenue management systems have kept up with changing business models, the focus has always been on optimizing the base product. With the rising share and importance of ancillary revenue to airline's bottom line, we saw an opportunity to bring dynamic pricing to ancillaries through collaboration with this airline.

Most airlines rely on both indirect (e.g. Global Distribution System) and direct channels (e.g. airline mobile app, dotcom) to distribute their products and services. Historically, indirect channels have lagged behind direct channels in adapting new distribution and pricing strategies. Most indirect channels now support branded fares, but à-la-carte ancillary sales still predominantly take place on the direct channel or at the airport. Given how airlines distribute and update prices in the in-direct channel, it is impossible to apply personalized dynamic pricing strategy on indirect channels. On the other hand, for direct channels, airlines have complete control on the price of à-la-carte ancillaries. For this reason, our focus for this project has been on personalized pricing of ancillaries in direct channels.

When it comes to pricing of à-la-carte ancillaries, most airlines use a static price or a simple rule-based engine for adjusting the price based on supply. In order to curate the rules, airlines rely on historical sales data to segment their customers based on the population's willingness to pay for an ancillary. Airlines then translate these segments to rules based on criteria such as departure date, days to departure, and market (origin - destination pair). They then typically observe what the competition is doing and align their own ancillary price to their competition for each segment.

Our Motivation

Our key motivation is to combine the aspects of dynamic pricing and personalization of ancillaries to achieve improved revenues; using the very rich context available in a live shop, to provide better predictions and recommendations. As no two customers are same, and no trips that a customer takes are the same, we aim to price the ancillary with respect to the specific context or shopping session. Instead of pricing an ancillary for a segment of passengers who are all assumed to be same, as is done in current revenue management work, we aim to contextualize the price to the time and point of sale. Such contextual pricing might be new to the airline industry, but has been consistently and effectively used in the past by companies like Amazon (Vermeulen and Bezos 2010), Uber (Chen 2016) and Airbnb (Ye et al. 2018). Moreover we aim to arrive at pricing strategies that are win-win from both the perspectives of the customers and the airline.

We categorize the rich "context" from a live shopping session as follows: (1) Customer context - offers, clicks, and selections that the customer has made in the session so far; (2) Trip context in terms of price of various flying options, departure time/date, length of stay and advanced purchase; (3) Supply vs. Demand context in terms of the current availability of the ancillary and past sales. The other opportunity we observe is to move the optimization of the pricing strategy to 'the edge'. Today, most pricing strategies are optimized offline using advanced analytical techniques and the devised optimal strategy is translated to a business rule, which is periodically updated to the various points of sale (distribution channels). Instead of packaging an "optimized business rule", we wanted to investigate if we can package powerful algorithms directly and deploy them on the edge at the airline.

We build upon the very promising nature of machine learning (ML) algorithms, which unlike business rules, are not limited by their capacity to handle complex logic. This ML algorithm(s) takes advantage of the rich context available at the time of the shop to make its predictions. The three elements of context discussed above are fed into the algorithm for predicting whether the customer will buy a given ancillary at a given price, and optimizing for the price which should be offered to the customer. The ML algorithm can also learn based on real-time feedback received from the channel.

Related Work

Despite an initial mixed response, many legacy airlines have recently adopted the strategy of unbundling, i.e. separating the main product into primary (right-to-fly) and ancillary products (bags, meals, etc.) for more customer flexibility (Garrow et al. 2012, Tuzovic et al. 2014, O'Connell and Warnock-Smith 2013).

Current state-of-the-practice on ancillary pricing at airlines leverages human-curated price lists, where analysts hand-pick prices based on their domain knowledge. This approach suffers from shortcomings of (i) humans being unable to react quickly to evolving customer preferences and market conditions, (ii) inefficiency and resource-intensiveness of continuous monitoring, (iii) inability to detect and control for humans' dynamic biases. These factors can be effectively mitigated by the use of computational models which can later be vetted by an analyst.

Ancillary purchasing behavior has been studied in the academic literature. Bockelie and Belobaba (2017) modelled ancillary product purchase, and compared the difference in behavior of customers who purchase ancillaries sequentially as opposed to simultaneously. More fundamentally, the principles of economics suggest that increased competition between airlines on ancillary products will increase equilibrium profits. Cui et al. (2016) demonstrated that an airline's ability to discriminate between customer characteristics greatly impacts its profits. More broadly, theoretical behavior models (Kahneman and Tversky 1979, Gabaix and Laibson 2006, Shulman and Geng 2013) can also be used to model customer behavior and choices, and their applicability in the context of airline ancillaries is poorly understood; but has recently seen renewed interest (Allon et al. 2011).

Contribution

The following quote summarizes our primary contribution to our partner Galactic Air's business metrics (revenue numbers correspond to ancillary sales from the airline's online booking engine between September 2018 - April 2019):

"*Deepair's personalized pricing system was excellent and improved ancillary revenue per offer (and thus total ancillary revenue) by 25% and ancillary conversion rate by 15%*" – Airline Group Online Transformation Manager.

The impact of the above increase in revenues is understood in the industry context through a statement in the presentation of a Boeing representative, "increasing revenue by 1% is equivalent to decreasing maintenance costs by 13%" (Boeing 2013).

Through this work and ongoing work with Galactic Air, we demonstrate that practical and fundamental changes can be brought to an airline's revenue management practices. We achieve this through a combination of new pricing algorithms, personalization and on-the-edge computation. The airline collaboration, described in this paper, started with pricing of ancillaries. Aimed at tailoring the ancillary experience for customers, this process helped us understand customers' willingness to pay which further led to increased ancillary revenues for the airline. By dynamically learning from customers' responses, our ML algorithms constantly learn about their needs and update the underlying pricing models.

Our models have been making pricing recommendations in the airline's booking systems since September 2018 (but have been on hiatus since the COVID-19 pandemic).

- From a data science perspective, we analyze multiple characteristics of historical shops and characterize customer features that seem to have an impact on ancillary revenue. We find that the following features had significant impact on purchase: date and time of departure, duration between booking and departure, market characteristics, length of stay, group size and base ticket price.
- From a mathematical modeling perspective, we present multiple types of ML algorithms for dynamic pricing while learning from customer behavior on-the-go. The first set of two models uses a two-step approach: customer purchase probability prediction followed by a classical pricing model to maximize expected revenue. The second type of model combines the two steps into a single end-to-end model through a neural network, thus resulting in enhanced performance due to lower error.
- To ensure true applicability, we first compare the models in an offline setting to measure multiple metrics of interest such as AUC, Regret Score, Price Decrease Recall and Price Decrease Precision; to estimate the revenue lost from pricing too low or too high. Through this offline evaluation, we choose the models that are performing well and deploy them for online testing. Given the novelty of this pricing approach, we carefully planned the online roll-out. We initially deployed the approach to one channel and one product and for 5% of customer traffic. Upon gaining data, insights and confidence in the approach, the airline rolled out the model to 70% of all traffic. In all, this deployment demonstrated improved revenue per offer (and thus total revenue) by 25% and offer conversion rate by 15%.

Data

We analyzed data that has been historically collected on Galactic Air's website. We classify each customer session (visit) into three broad categories: (i) the session's input parameters (such as origin, destination and time of departure), (ii) details on all of the offers made to the customer based on the input parameters, (iii) the user response click-stream of what offers were clicked and/or selected. We saved these offers and user data for every session (both buying and non-buying), for six months prior to the start of the study and during the entire study itself.

There are many features in the collected data that contain valuable information about the price sensitivity of each user session. Classical approaches involved humans manually looking through the data and trying to spot patterns that would help guide a static price based on economic factors. More advanced approaches have used price points for different sub-markets or other categories, such as seasonal pricing. Here, we are interested in building models that learn relationships between these factors and customer choice or behavior without having to define these interactions *a priori*. There are many features that we believe would be useful for the models. We outline a few important features that we identified before running the experiments. They were used alongside other features, such as group size, advanced purchase, and base ticket price, all equal weighted for model initialization. Other features were included but cannot be discussed due to confidentiality restrictions.

Time

There are two types of time-related factors that we will highlight as having particular influence on prediction: 1) *Date and time of departure*: We observed a dependence of time-of-day and seasonal time on buying behavior. Itineraries beginning on particular days and times are in higher demand; due to factors such as convenience, connectivity, more attractive availability of alternative connections, holidays, and special events. 2) *Duration between booking and departure*: Price-sensitive customers tend to purchase their tickets well in advance.

Markets

Galactic Air has a large route network serving a variety of markets. A market describes an airport pair or city pair consisting of an origin and destination, which may be connected by intermediate airport(s). The markets are diverse with some having non-stop itineraries while others contain many connections via the airline's hub airports. Different markets appeal to different customers and itinerary types (e.g. business and leisure). Figure 1 visualizes the different markets based on their demand for ancillary services. It is clear that ancillary demand varies greatly between markets. This suggests that market information provides useful context for predicting if a customer will purchase an ancillary. Although it is clear that the market influences probability of purchase, it is not clear what the optimal pricing strategy is, particularly when non-linear effects of other features also influence customers.

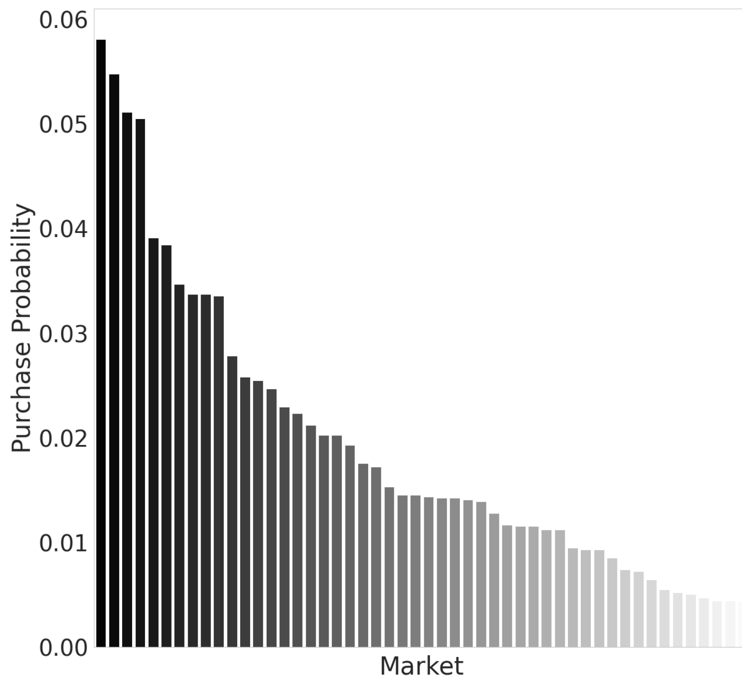

Figure 1 **Market clusters ordered in decreasing probability of ancillary purchase, showing significance of 'market' as a feature, with different markets exhibit different average probabilities of purchase.**

Visualization of historical purchases with t-SNE produces the embedding shown in Figure 2. Figure 2 shows the instances where the ancillary is purchased (black cross) and not purchased (grey circle). This indicates the presence of regions in the embedding that have varying purchase probabilities, with some "all-grey-circle" (no ancillary purchase) regions present. Note that even in the most successful regions, ancillary purchase fraction is small.

Length of Stay

Length of stay (LOS) is the number of days between outbound and return flights. This applies only to bookings that are round-trip. If the itinerary does not have a return, we consider LOS to be 0. Figure 3 shows the kernel density function for LOS for two types of itineraries: when an ancillary was purchased (dashed orange line), and for all bookings (solid green line). The frequencies vary considerably between the LOS of 0.0 to 0.3, suggesting that customers prefer ancillaries for medium length trips. This deviation between the two highlights the importance of LOS on ancillary demand.

Data Architecture

In Table 1 we describe the different data categories, which are all present in Galactic Air's database. The architecture of our approach connects the data to the pricing models and

Figure 2 t-SNE embedding plot of historical purchases, marked based on purchase (grey circle) and non-purchase (black cross) of individual shopping sessions, shows that ancillary purchase fraction is small in general, but we observe higher purchase fractions in some regions of the embedding than others.

then to the A/B testing framework, as described in Figure 4. 'Transaction Storage', 'Big Data Pipeline' and 'Data Cluster' are the data layer. These connect to Deepair's model parameter tuning and pricing engines (described in section 'Pricing Models') through the online training and deployment pipeline. We then send the pricing solutions to the A/B testing framework which chooses between the Deepair pricing engine and the human-curated pricing engine to the customer touch point, which is through Galactic Air's website. As needed, we can also incorporate additional touchpoints such as the app or other booking engines.

Pricing Models

We set out to achieve two objectives at our airline partner: 1) maximize expected revenue per booking offer, and 2) maximize conversion rate of bookings. Concurrent focus on both helps maintain a balance between sustainable business for the airline and improved customer satisfaction via booking offers more likely to be accepted by the customers. Maximizing revenue alone can result in the model finding a sub-optimal solution where extremely high-priced offers result in overall satisfactory revenue. Such a system would be

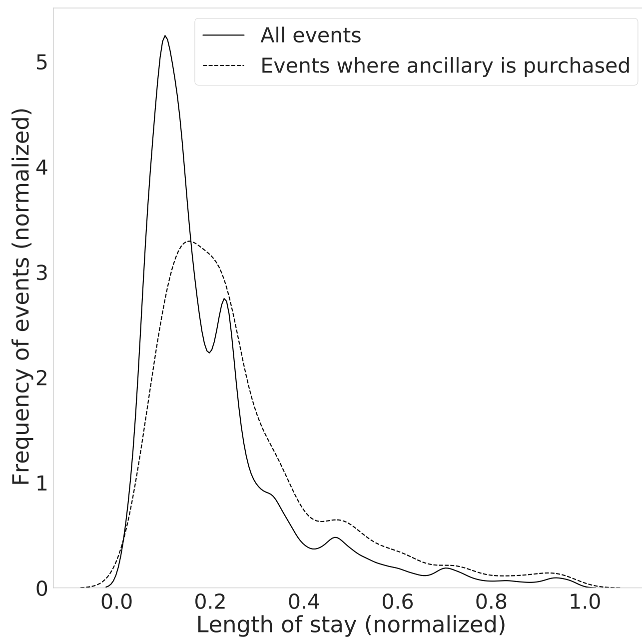

Figure 3 **The Length of Stay (LOS) signal from Kernel Density Estimation (KDE) for sessions where an ancillary was purchased, shows that passengers prefer purchasing ancillaries for medium length trips**

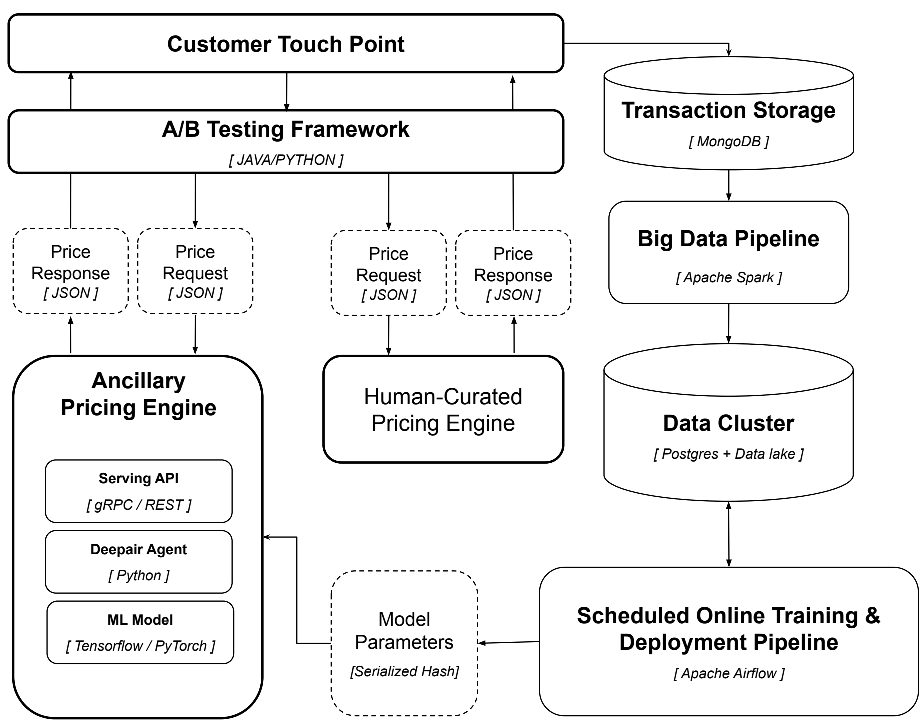

Figure 4 **Architecture depicting flow of data in our implementation**

Fields	Types
Origin	Categorical
Destination	Categorical
Departure Time	Datetime
Arrival Time	Datetime
Request Time	Datetime
Passenger Type (Adult/child)	Categorical
Advanced Purchase	Time delta
Length of Stay	Time delta
Clicked/Selected offers	Categorical
Offered products	Categorical
Market Volume	Numerical
Market Popularity	Numerical
Flight Historic Events	Hashed

Table 1 Data components used in the pricing models

over-reliant on a small subset of customers who are willing to accept higher-priced offers. However, expected customer satisfaction would be poor as the majority of the interested customers would decline the high price. Therefore, it is critical to understand and match booking prices to the customer's desire of purchasing a product in that specific booking session. Here, we use the customer's purchase probability as a metric to understand the underlying needs and offer targeted discounted prices to the ones associated with low purchase probability. This mechanism allows us to capture otherwise missed revenue, and simultaneously results in an enhanced customer experience.

It is important to note here that we restrict the price ranges we consider for our ML models to be lower than the human-curated prices for each customer group. Thus, from a customer perspective, if an ancillary is chosen, the customer always pays less than the historically static human-curated price. To predict a customer's purchase probability for a given product, we begin with a large database containing previous user sessions - that includes a variety of factors useful for prediction along with a label that encodes whether the customer purchased the product or not. We describe this in the Data section. We now discuss the models we develop for price recommendations to customers.

The three models we discuss for price recommendation fall into one of the two pricing strategy classes mentioned below. The first type of approach uses a two-stage model – beginning with capturing individual demand to predict the individual's purchase probability as a function of price; and then followed by a mapping of that probability to a final offered price. We use this in Model 1 and Model 2, described in detail below. The second type of approach directly predicts the final offered price from the input data. It is also referred to as end-to-end learning in the machine learning community. We use this approach in Model 3. An overview of the three model architectures is provided in Figure 5.

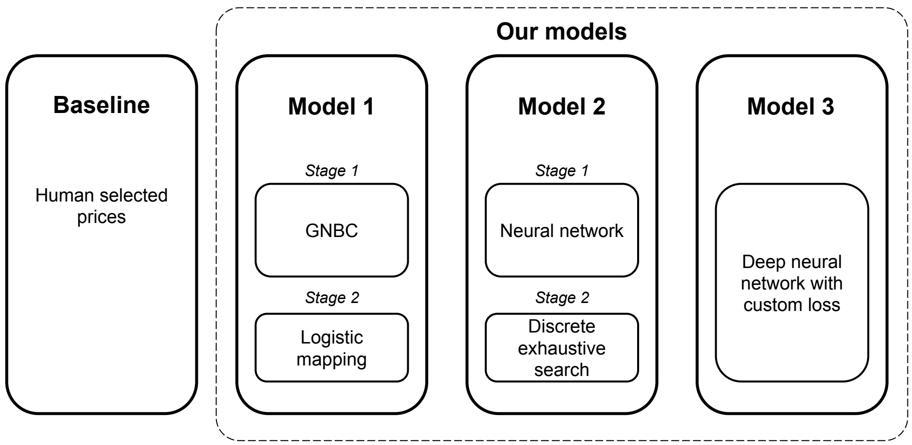

Figure 5 **Overview of the pricing models - Model 1: Two-stage model with logistic mapping, Model 2: Two-stage model with Discrete Exhaustive Search, and Model 3: Single-stage end-to-end deep neural network; all compared against the baseline, which are prices selected by human analysts.**

Models 2 and 3 use neural networks at their core, though in different ways. Model 2 uses a neural network to predict customer purchase probability in the first step (and then a separate search process in the second stage to find the best price) whereas Model 3 uses a neural network to directly provide a recommended price. To provide background, a neural network is a set of sequentially connected layers, where each layer has a set of units (or neurons) that operate on the outputs of the previous layer. In most cases each unit performs a weighted sum of its corresponding inputs from the previous layer. Parameters of the neural network (weights of this sum) are learned by a backpropagation algorithm (Parker 1985, Werbos 1974). The sum is then propagated through a non-linear activation function before being used in the next layer. Concatenating multiple layers one after another results in a *deep* neural network which possesses the power to learn a rich representation of the data relevant for the task.

In both Models 1 and 2, having a probability of purchase is the first part of the pipeline to our objective of producing an optimal price. The second stage is a mapping from purchase probability to a final recommended price. In Model 3, the neural network is directly optimized for the price to be finally offered to the customer; thereby integrating the willingness to pay within the neural net. Because deep neural networks are remarkably robust to noise and do not require processing (or feature-engineering) of the input data (LeCun et al. 2015), this can significantly reduce the work of data scientists and streamline much of the analysis and learning from data.

Model 1: Two-stage Model with Logistic Mapping for Pricing

In the first step, this model uses a Gaussian Naive Bayes with Clustered features (GNBC) model for predicting customer purchase probability given input features. GNBC is a probabilistic prediction model that works under the (prior) assumption that the features are independent. This often works well with real-world data that is of different types (Rish et al. 2001). The second step is a calibrated logistic mapping that optimizes revenue and conversion rate. Figure 6 shows a schematic of the method to map purchase probability to recommended price. Details are in the Appendix on Pricing Models - Logistic Mapping.

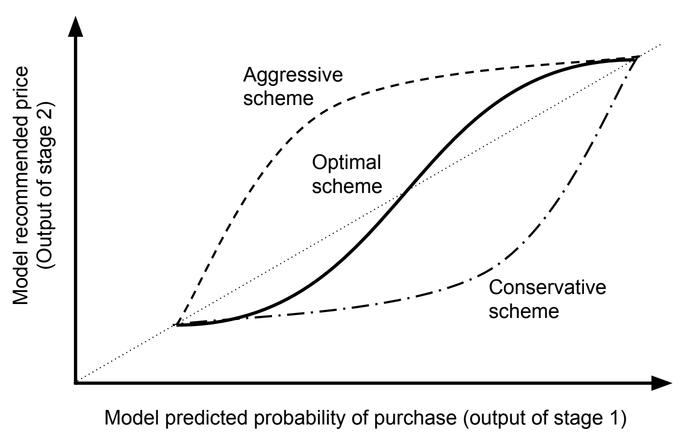

Figure 6 **Schematic of the logistic mapping model mapping probability of purchase to a recommended price, illustrating three options: aggressive pricing, conservative pricing and optimal pricing**

Model 2: Two-stage Model with Discrete Exhaustive Search (DES) for pricing

Our second model is also a two-stage model that requires that the purchase probability be first estimated. In this case we learn the customer purchase probability using a neural network trained on the same set of inputs as the previous model. Neural networks are a series of connected layers that learn an abstract representation of the input data. With an appropriate loss function, a network can be trained to learn a useful representation that minimizes the loss and performs the intended task (LeCun et al. 2015). In the second step that maps purchase probability to price, this model performs an exhaustive search over a set of admissible prices. Specifically, under reasonable (unimodality or convexity) assumptions, as seen in Figure 7, we efficiently perform exhaustive search over a small set of discrete prices that are within a pricing range that is chosen according to expert opinion. In our data we made the assumption that the revenue as a function of offer price is unimodal. We provide further details of the model in the Appendix on Pricing Models - Discrete Exhaustive Search.

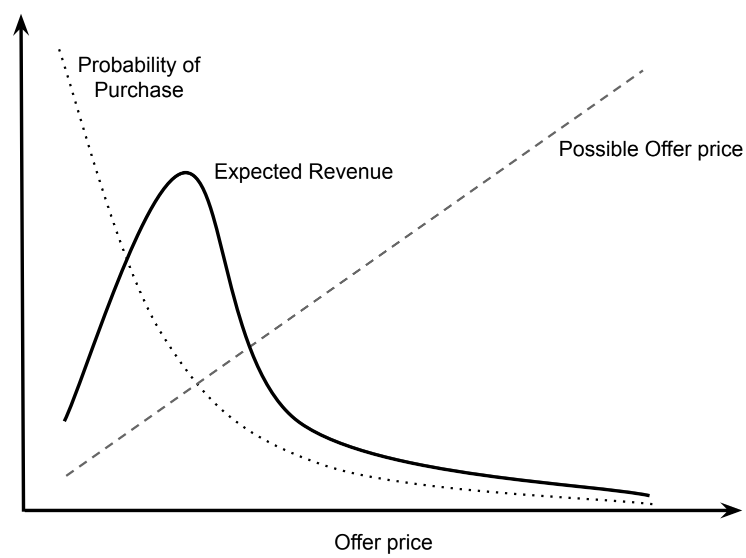

Figure 7 **Discrete exhaustive search (Model 2) uses predicted probability of purchase (output of stage 1, dotted line), and the information about allowed offer price ranges (dashed line) to calculate the expected revenue (black line); and chooses the price that maximizes revenue.**

Model 3: End-to-end Deep Neural Network

The third type of pricing model does not predict the purchase probability of the ancillary at various prices as an intermediate step. Instead of first predicting purchase-probability and then mapping that to a price point, this model integrates the two steps and optimizes for offered price directly from the input features in a single-step. This is also known as end-to-end learning. This strategy has a trade-off in that it is capable of learning a richer set of relationships between customer behavior and price because of its integrated nature; however, it is potentially less interpretable.

Here, we use a neural network with two layers. Training this end-to-end model to predict the optimal price is not as straightforward as the previous two models, where we predicted purchase probability. For the optimal price we do not have a complete "ground truth" signal, because historical data only reveals a yes/no answer on whether the customer purchases the ancillary at the offered price. If the customer did purchase, we know that the willingness to pay could have been higher and vice versa if they did not purchase. For example, say a customer purchased a product for \$10, then that same customer might have been *willing* to pay \$11. In this model, we are not interested in predicting the price at which something was purchased in historical user sessions, but rather the *optimal* price. However, this is not reflected in the data and predicting the purchased price directly will not result in a model that fits the task of finding the optimal price. This issue makes this model different from Model 1 and Model 2. To alleviate this, we designed a custom loss function to account for not only the regret of pricing too low but also an additional penalty for pricing too high. Our formulation builds upon concepts introduced by Ye et al. (2018) and Shukla et al. (2019). Mathematical details of this model are in the Appendix on Pricing Models - End-to-end Deep Neural Network.

Essentially, this custom loss trains the model to learn in a manner that is considered logical based on previous experiences. For example, say a previous customer had *not* purchased a product for \$10 then, logically, the model should not price at \$10 or above for another customer with similar features. Say a model outputs \$13 for such a customer during training, then loss would be high and penalize the model to reduce the price.

Our proposed loss function is inspired by the ϵ-insensitive loss used in Support Vector Regression (SVR) by Drucker et al. (1997), which is ideal to determine upper and lower bounds for willingness to pay. These bounds provide a richer reward signal for multiple models to learn willingness to pay. For example, assume model A recommends a price of \$10 for a customer which is accepted and model A gets a binary reward. A different model, B, can make a recommendation for the same instance (one that is never shown to the customer). Say that recommendation is \$8, we assume that will result in a sale but we also have *additional* information that the customer would have purchased the ancillary at the lower price. Using this information, model B can be trained with a richer reward signal. Like SVR, with significant data, our proposed custom loss function converges to a tighter lower and upper bounds (i.e. closer to actual willingness to pay) over the entire population for suitable hyper parameters. Hence we do not estimate the true willingness to pay, but a range within which it could belong.

Offline Experiments

Before deploying the models in Section into operation for real-time pricing, we first prepare and validate them in three phases. First, we train the models on collected data from previous user sessions. Second, we tune the models' hyperparameters to optimize their performance. Third, we validate the tuned model performance to show a measurable improvement over the existing pricing strategy in offline tests to provide evidence to our airline partner prior to online implementation. Our offline scoring metrics were selected to gauge performance with respect to both economic viability and prediction accuracy.

Predictor performance

All of the models rely on accurate prediction of the customer's probability to purchase a product. Specifically, this subsection deals with the first steps of Models 1 and 2, which explicitly predict customers' probability of purchase (but does not recommend a price). For these models, we can measure their performance directly using the ground truth labels on historical data. We use the *area under the receiver operating characteristic curve* (called AUC-ROC or AUC in brief) as a measure of how well our models are performing in predicting whether the session resulted in a sale or not (binary classification). The training dataset is quite sparse, that is, the historical data has many more instances of people not buying the product than those buying (as also seen in Figure **??**). This creates an imbalanced ratio of labels that has to be accounted for obtaining an unbiased measure (if 90% of sessions end without a sale, then a model that predicts all sessions as non-sale would have have 90% accuracy). Comparing the AUC gives a more balanced insight into the predictor – an AUC of 1 represents a perfect predictor while a model that makes random predictions would achieve an AUC of 0.5.

We started by creating a baseline that represented the current state of the airline industry. This model was a Gaussian Naive Bayes with Clustering (GNBC) predictor. We then compared this baseline with a number of more advanced models, namely Gaussian Naive Bayes (GNB), Random Forest (RF) and Deep Neural Network (DNN) predictor, where all outperformed the baseline, as seen in Table 2. DNN achieved a 33% improvement in the AUC score compared to baseline and displayed the most dominant performance overall.

Table 2 AUC score of models on datasets. Deep neural network (DNN) outperforms the other models: Gaussian Naive Bayes (GNB), GNB with Clustering (GNBC) and Random Forests (RF) on three different dataset configurations A, B and C.

Dataset	GNBC	GNB	RF	DNN
A	0.54	0.60	0.66	**0.68**
B	0.53	0.62	0.68	**0.70**
C	0.57	0.63	0.66	**0.77**

The three datasets presented - A, B and C - comprised $41,000$, $50,000$ and $72,000$ sessions respectively. WE introduce variation due to seasonal changes within the datasets by random sampling. For example, number of holiday seasons present in datasets: $C > B > A$. DNN subsequently shows that it is best capable of leveraging larger quantities of data effectively, with its relative performance compared to other models increasing as dataset size is increased.

Pricing performance

This subsection applies to all models that provide a final pricing recommendation, namely Models 1 and 2 (after the second step of recommending a price has been executed), and Model 3 (which directly recommends a price). For pricing performance metrics, we use the *Regret Score* (RS), *Price Decrease Recall* (PDR), and *Price Decrease Precision* (PDP), inspired by Ye et al. (2018).

Intuitively, the Regret Score captures the missed opportunities when recommended price is below the (historical) true purchase price. For example, if a customer purchased an item for \$10 and a model later recommended a price of \$8, then the regret score for that session would be 0.2. In the case where the model predicted a price higher than \$10, the regret would be 0.

PDR is the proportion of non-purchased historical sessions where the model recommended price is lower than the true purchase price. If PDR is 1, it indicates that the model's recommended price is always lower than the historical price in the data for those sessions which didn't end in a sale - this would be a desirable result. Conversely, PDR is 0 when every new price is higher than the historical price for non-purchased sessions: an example would be a model that recommends a price of $13 for a session where the product was not purchased at $10. PDP is the proportion of sessions where, given that the model's recommended price is lower than the true purchase price, the the product was not purchased. For the mathematical definitions of these metrics, see the Appendix on Performance Metrics.

Finally, we use a *Price Decrease F1* (PDF1) score to evaluate the balance between PDR and PDP. It is inspired by the traditional F1 score that measures the trade-off between precision and recall.

Table 3 Comparison of scores for different models in offline experiments for Dataset C**.** Lower Regret score (RS); and higher Price Decrease Recall (PDR), Price Decrease Precision (PDP) and Price Decrease F1 (PDF1) indicate better performance.

Scores	Logistic mapping (Model 1)	Discrete exhaustive search (Model 2)	End-to-end NN (Model 3)
RS	0.07	0.38	0.07
PDR	0.64	0.63	0.83
PDP	0.93	0.93	0.92
PDF1	0.76	0.75	**0.87**

In Table 3, we provide an overview of the offline pricing performance. This allows us to gauge each model's behavior in a controlled offline setting, on real historical booking events. Model 3, a DNN with our custom loss, performs significantly better than the two other models in terms of PDR. However, because the PDP scores for all models are comparable, PDF1 score which is a function of both PDR and PDP gives an advantage to Model 3. With this discrepancy between Model 3 and other models, we hypothesized that Model 3 was recommending low prices for non-purchased sessions.

Given that we were providing dynamic pricing in one direction from the historical human-curated price (only discounts are allowed) we treated RS and PDR to be the most important metrics. In these priority metrics, Model 3 performed best, suggesting that it would perform best in further online experiments. This also suggests that the end-to-end neural network (Model 3) is able to learn the relationship between session context and optimal price to the degree required for recommending prices that improve the metrics. However, note that the neural network predictor that performed well in the previous prediction task (Section), and is used as the first step in Model 2, does not perform best. We overcame its limitation by extending the neural network to become an end-to-end model that, which when coupled with our custom loss, performs best (Model 3).

Real-world Implementation and Results

After measuring the performance of the models in Section using the offline experimentation and validation (Section), we deployed them in live testing and measured their performance in the real-world.

Here we quote the Head of Digital Transformation at our airline partner: *"Deepair's modeling frameworks were easy to understand, easy to implement and had a plug-and-play type feature allowing for easy testing and easy scaling from a small subset of customers to larger and larger customer groups."*

Like many other digital transformation projects, this project was also approached with the three P's in mind - Product, People, and Process. We detail each of these below.

Product

The airline partner put forward some key features they wanted us to incorporate in the delivered product.

One of these was *context-sensitive pricing at the edge*. Most pricing solutions are centralized - housed in a central reservation system. Such a framework doesn't incorporate context of the shop; thus, it was important to take the algorithm closer to the edge by making it highly responsive at a session level. As each booking channel captures different context (features on the website could be different than the mobile app), we had to deploy different algorithms catering to different features at various edges. Variations in the context could be: 1) Context changes based on the current position of the customer with respect to the journey. For example, context of ancillary sale during flight booking is quite different from the context post-booking. 2) Context variation by product (e.g. bags, seats, priority boarding, etc.) being sold. We decided to design the system such that each product uses a different algorithm.

In order to measure the effectiveness of algorithmic pricing, it was important to measure it in conjunction with the current pricing policy and systems. We leveraged the *A/B testing* tool already in use at the airline, and adapted it to test rule base pricing (human-curated) vs. algorithmic pricing (our models in Section). Continuously improved algorithms with different hyper-parameters demanded a mechanism for bridging the gap between the offline and online settings. Hence, we built a key feature into the product, thereby allowing for the ability to run many algorithms in parallel.

As ancillary revenue is a major contributor to an airline's growth, there are specific departments and personnel responsible for the business outcomes. Hence, it was important to provide the airline analysts and management with the desired tools for gauging how the ancillary business is performing and subsequently managing levers to control the algorithm's behavior. The analysts are responsible for setting a hard-threshold on the minimum and maximum price possible for each ancillary within which the AI could function. They also turn or off dynamic pricing in certain markets, products, or sessions, in order to ensure that the airline is in compliance with the federal authority regulations.

Finally, we put a lot of design thinking into the architecture of the solution. Contextual pricing is a critical system for the airline. Our algorithms are expected to provide sub-second response, hundred of times every second depending on the load at the airline's website. Responsiveness is an important feature of the product and goes a long way in inspiring confidence among the IT/e-commerce departments on the quality of price being served to the customers in each booking request.

People

We made sure that the key stakeholders from across the airline organization were aware of the scope of this project and the impact from the dynamic pricing of ancillaries.

The head of the ancillary revenue department, typically reporting into the Chief Commercial Officer, was the key sponsor of this project. Previously, there were ancillary pricing analysts responsible for managing the business rules which control the pricing. We were going to empower these analysts with new tools, further allowing them to focus on more strategic aspects of their job and delegate the minute price variations to the algorithm. We realized it was important to build the right levers and visibility into the algorithm so that the analysts could use the tool with confidence. The ancillary revenue department still had control on what the dynamic pricing system was allowed to do.

A range of other departments were involved in setting up the live system. The Data Analytics department were heavily involved to make sure the right data is captured and sent to the algorithms for training. At most airlines, a majority of the ancillary sales occur either through the digital platform or at the airport. For this work, we were fully focused on the airlines' internet booking engine (also known as *Dotcom*). Through this we were supported by the e-Commerce Platform team as the nature of the product was such that it required tight integration with the e-commerce platform. The dynamic price recommended by the algorithm was ultimately presented to the customer (end user) through this platform. The marketing team played a key role in rolling out any dynamic pricing project. If not communicated properly, a dynamic pricing roll-out may backfire and cause customer distress. This team was involved from the beginning of the project to ensure the right means of communication and the right verbiage is used while rolling out dynamic pricing. The sales department worked on targets as any small change to the pricing schemes would directly impact their ability to meet their predefined targets. So, it was important to make sure that all the configurations were approved by the sales team.

The airport is another major channel for ancillary sales. Most airlines want travelers to purchase ancillaries before they arrive at the airport, in order to avoid long waits and operational delays. This airline made a policy decision to reflect that the dynamic price available on the digital channel is always cheaper (substantially in some cases). We informed airport staff of these policy decisions – so that they are able to answer questions from customers on any price discrepancies. In order to implement dynamic pricing, changes had to be made to the airline's back-office processes such as revenue accounting. This was handled by the finance department. The legal department was involved to make sure that the airline is in compliance with regulations and the terms & conditions set between the airline and their customers. Finally, customer support team was informed about these changes – so that they are able to explain to the traveler why the price was dynamically changing and how the price is determined.

Process

The third key "P" that led to a successful deployment is the process that was established around the dynamic pricing system. Critically, we introduced two metrics, one airline-oriented and one customer-oriented, for keeping track of business goals. Our goal was to make sure we improve on both metrics. *Revenue per offer (RPO)* measures the economic benefit of dynamic pricing to the airline. If this metric increases, it indicates that the airline is making more money from the product. It also measures the overall effectiveness of finding the right price for every offer. If the price is too low, the airline is giving it away for a lower price (dilution) despite customer willingness to pay; and if the price is too high, the airline is turning down a sale (spoilage). *Orders per offer (Conversions)* indirectly measures the customer satisfaction or how 'fair' the customer perceives the price to be. If this metric is increasing, it indicates that we are able to successfully match/stay below the customer's willingness to pay.

Next, we wanted to make sure that RPO and Conversions are measured with the A/B testing module in place. In particular, we compare:

- Daily RPO and Conversions - human baseline vs. dynamically driven pricing.
- Daily RPO and Conversions for each live model.

Initial Deployment and Roll-out

In order to manage the business uncertainty and financial risks, we rolled out the system in phases: 1) We started with one-product, one-touch point, and a small group of customers. 2) Once we felt confident our models were improving on the baseline in terms of the RPO and Conversion, we extended dynamic pricing to a larger population of customers. 3) Once we found a model was statistically better than another model, the inferior model was be retired and replaced by the better one or a newer model. 4) We extend this to multiple products and multiple touch-points.

When we first started the project, Galactic Air, our partner airline (as is standard at airlines around the world) was using primarily static price points for ancillaries that rarely changed. This meant that the algorithms were unable to fully understand customer's price sensitivity to a product, as only one price point had been observed. To overcome this problem, we designated an "exploration" period at the beginning of the project - where we would offer random discounts to a subset of the population to collect data. The airline understood that random exploration would potentially result in temporary revenue loss, but this cost was offset by the value gain in additional information.

Figure 8 shows the overall deployment of the system. The customer touch-point (in this case the airlines online booking engine) connects to the Ancillary Pricing Engine (our models) through an A/B testing framework. The A/B testing framework controlled the

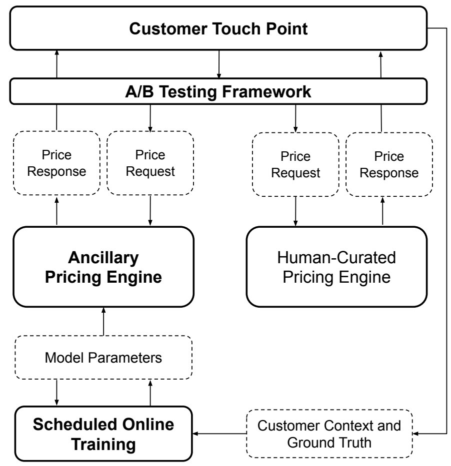

Figure 8 Deployment schematic of the Ancillary Pricing Engine (Dynamic pricing) and the Human Curated Pricing Engine in an A/B Testing Framework

percentage of traffic that was sent to our models vs. the current human-curated pricing engine.

In the initial phase 5% of the website traffic was directed to our models. When enough information had been collected in the exploration phase, we deployed online a selection of models that performed best in the offline tests. We compare three models of each model type (Model 1, 2 and 3). Note that in offline experiments and testing (Section), Model 3 significantly outperformed the other two model types, using offline metrics for comparison.

Figure 9 shows a comparison between Model 1 and Model 2, with the baseline included for reference. Both models outperformed the baseline on most days with Model 2 showing slightly higher variance. In online tests, Model 3 remained the best performing model, but its advantage was not as significant. Note that this discrepancy between offline and online performance is expected and is termed the “realism gap”. Estimating the effects of the realism gap helped us select models that we expected to perform better online. In this this phase we quickly improved our models and saw an improvement over the baseline pricing. With continued performance gains of our models, the airline decided to significantly increase the traffic share sent to our Ancillary Pricing Engine from 5% to 70%.

Extended deployment

The extended deployment phase describes the timeframe after which more than 70% of traffic was directed to our Ancillary Pricing Engine. In Figure 11, we see the total improvement of our models over the baseline (human-curated prices); representing 15% higher

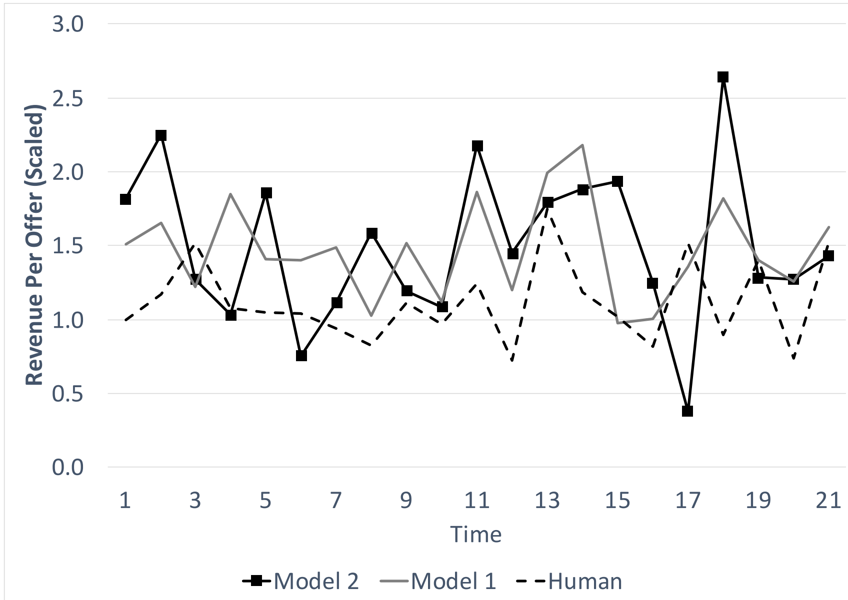

Figure 9 **A comparison of revenue per offer by model shows that Models 1 and 2 outperform human-curated prices (axes values are normalized to protect business sensitive information)**

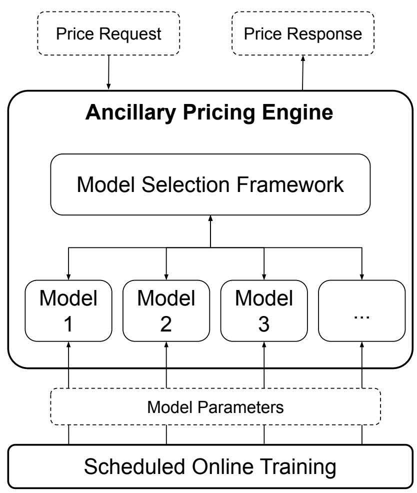

Figure 10 **Model Selection Framework in which multiple models are deployed and tested in parallel in an online setting**

conversion and 25% higher ancillary revenue per offer. This translates directly to week-on-week improved ancillary revenues of 25%; and improved revenues of 20% over the entire initial and extended phases when the traffic sent to these models was varied. With the experience learned from the initial deployment phase, our models dynamically capitalized on the information gained with the consistent 70% traffic and improved our models even further.

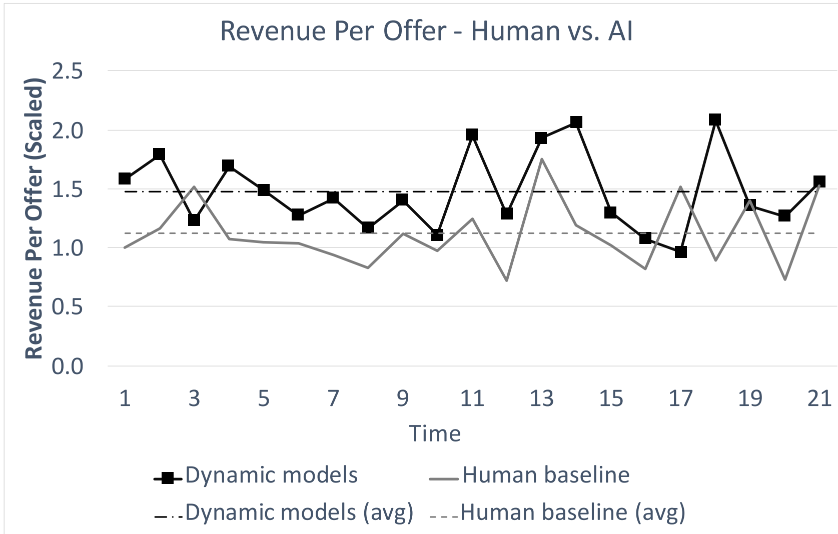

Figure 11 **Our online dynamic ancillary pricing models consistently and significantly outperformed the baseline in terms of revenue per offer made (axes values normalized to protect business sensitive information)**

With the A/B testing framework configured to divert 70% of the traffic to our models, we were able to get a much clearer signal of model performance. We were also able to improve the frequency of online training of our models. With enough traffic coming our way, we deployed many model configurations, hyper parameters and algorithms in parallel to test their relative performance. As shown in Figure 10, we implemented a model selection framework within the Ancillary Pricing Engine that would randomly divert traffic to various models. Every time a model showed promise in the offline metrics, we were able to test it in the online setting as well.

Implementation Experience

Booking systems in the airline industry were not originally designed to have ancillaries separately sold from the base product. Hence, we had to design new systems from the booking perspective as well as the customers' perspective. Once Galactic Air adopted a systematic design of new booking systems, along with deep data science perspectives as well as integration into user experience, we observed that overall, ancillaries contribute between 10 and 40% of revenue.

Airline customers' response: Historically, airline customers have been comfortable with the fact that dynamic pricing and revenue management are used extensively in the airline industry. Moreover, the adoption of these practices in applications such as ridesharing and hotel booking have also habituated customers in other industries to these practices. In this implementation, customers were informed up front that price of the ancillary could be lower than the standard fixed price that they have been seeing over the past years. That primes the customer to possibly see a discount when they get to the stage of purchasing an ancillary in the booking process. We expected that because they see a lower price, conversion rate will increase. As expected, we found an increase in ancillary demand – indeed, possibly a stimulation in demand as customers who might not have bought a bag previously are incentivized to purchase at the lower price.

We also find that there is no statistically significant change in the demand patterns due to dynamic ancillary pricing. Moreover, the demand that result in purchases of tickets has also not significantly changed, and there is no noticeable shift in the number of shops of tickets. Fundamentally, some of this is attributable to the fact that each ancillary by itself forms a small fraction of the overall ticket price; and is an add-on, not the base product. Dynamic pricing, therefore, does not impact the decision of a customers to be loyal to the airline.

Operational costs due to increased ancillaries: We specifically refer to bag purchases here, as they could increase the payload carried by the aircraft and hence the operational costs due to fuel burn. To address this, we ensured that the unit fuel cost of an extra bag was always kept lower than the least price offered to the customers for a bag. It turned out that this incremental cost of adding bags was quite low, allowing for a cap on the total number of bags for a flight. Moreover, there is a cost advantage to customers buying bags a priori compared to the airport, as some customers preferred to eliminate (free) cabin bags due to the low cost and added convenience of check-in bags. Increased bag check-ins also resulted in a smoother boarding process, and reduced delays at departure, improving on-time performance. Thus there were some unintended positive consequences of passengers making earlier decisions to purchase bags.

Adaptation by marketing personnel: Marketing personnel adapted to the implementation of the new dynamic pricing strategy through an intensive internal messaging exercise. This included a description of the price to be in a range instead of a fixed value (as it is when prepurchased at an airport). Because our pricing strategies were only adopted on the dotcom channel, an internal roadshow was done to educate internal sales teams and call centers about this new pricing approach. Beyond an announcement to customers that bad prices were being discounted compared to the fixed price, additional customer marketing was not undertaken as part of the target was also to discover and learn customer behavior. Moreover, as this could be perceived as part of an experiment, GDPR compliance required customer consent. Hence, only customers who are registered members of the airline club were involved in this pricing policy as they have already accepted such terms and conditions implicitly and viewed the new pricing as part of the several discounts that such customers are entitled to.

Linking of systems for dynamic pricing: Legacy airline booking systems cannot accommodate the novel pricing mechanisms proposed. Hence, we adopted an innovative application of promotion codes in the background of the pricing system, so that a dynamic price could be offered, instead of changing the booking system itself. Following the initial application and during the COVID-19 pandemic, the booking system is being overhauled so that dynamic pricing can be an integral part of the booking system once operations resume as normal. This modification is also compliant with the NDC (New Distribution Capability), a new standard proposed by the IATA (International Air Transport Association), which enables dynamic pricing and is required to be adopted by airlines in the next five years.

To enable the promotion code technique to be adopted, the airline linked the e-commerce and dotcom booking systems, and changes were made to the clickstream data collection to enable learning of customer behavior. This was upgraded to send information in real-time to the learning modules, which in turn was linked to the A/B testing approach that the e-commerce system links to. The A/B testing approach was designed to direct part of the traffic randomly to Deepair's pricing platform and partly to the airline's legacy pricing platform. The fraction of traffic direction was also set up in a manner that could change over time as performance was observed.

Additionally, the airline also integrated the e-commerce system to the inventory system to set a limit on the number of ancillaries sold, and to indicate in real-time the number that can still be sold.

Impact

We compared the week-to-week improvements in revenues from our model using the A/B testing platform. When the traffic in the A/B testing setting was directed equally to the

Deepair pricing model and the human curated pricing at Galactic air, the deep learning model increases *ancillary* revenues by 20% compared to the standard pricing models, and by 25% when 70% of the traffic was directed to Deepair's pricing models. We found that the 20% ancillary revenue increase (over varying fractions directed to the human-curated and Deepair pricing models) was consistent when compared against the corresponding weeks in the previous year.

Conclusion

While many industries in the modern context have been able to take advantage of personalized dynamic pricing, this has proved extremely difficult in the airline industry due to its legacy booking systems and the sheer complexity and size of the factors and markets considered. We present data, methods and implementation-based evidence to show that this is possible in the airline industry. In fact, it can be done efficiently from our experience.

A machine learning based personalized dynamic pricing system, like the one we propose in this paper, is a mission critical tool for any airline – especially in times of uncertainty such as the COVID-19 pandemic. As customers start traveling again and airlines resume operations, airlines are going to fiercely compete to recapture market share. Any airline pricing analyst looking to capture market share would be leveraging a discounting strategy. Airlines that leverage smart, targeted and personalized discounting strategies, such as the ones we propose in this paper, can avoid revenue dilution and emerge as winners in the post-pandemic era. Moreover, customer behavior is likely to change due to this pandemic, so a machine learning algorithm that observes and adapts to changing customer behavior is going to be more efficient than algorithms that solely rely on historical data.

Through this work we also aim to show that a personalized dynamic pricing system creates a win-win situation for both the airline and its customers. While the airline enjoys higher revenue and higher sales, customers enjoy a sense of fairness because the price is sensitive to their needs and their willingness to pay. We demonstrate this by the facts of not only increasing revenue per offer, but also the conversion rate (which normally indicates that the customers liked the price they saw). In other words, the algorithm was able to recommend a price that was low enough to convert and avoid revenue spoilage, but high enough to sell cheap and avoid revenue dilution. This balance is hard to achieve through a human-curated price.

We presented our work on pricing ancillaries at a partner airline where our models led to 15% higher conversion and 25% more revenue per offer than the airline's contemporary (human-curated) pricing policy through a real-world implementation. By using reliable evaluation metrics that correlate well with business impact, we were able to partly develop our models offline using historical data before testing them online. Using A/B testing, we could improve our models' performance consistently by retiring sub-optimal models. A major feature of our system is its ability to react to changes in customer behavior. Although we have greatly reduced the realism gap, offline experiments can never fully replace online experience as changes in customer behavior will be delayed across the offline data.

Our work demonstrates the promise of improved business value through precise, accurate, and continuously updated models for customer demand and associated price sensitivity for ancillaries. In the future, we plan to implement online multi-armed bandit approaches that could be used to adaptively learn a better routing scheme to generate higher revenue per offer and better conversions compared to a randomly selected model, when dealing with a bucket (ensemble) of models. Furthermore, we can extend such a model selection framework to contextual multi-armed bandits with caution and concurrence, to optimize for different business metrics. We could also model customers' elasticity towards an offered ancillary to enable the simulator environment to model ground truth more closely. Finally, although this paper focuses on pricing for a single ancillary, the framework presented here can be applied to multiple ancillaries concurrently (a multi-agent system) and base ticket pricing.

Acknowledgments

Pricing Models

Logistic Mapping

In our model 1, once the ancillary purchase probability is predicted, we use a logistic function to recommend a price based on the predicted probability. The intuition behind using logistic mapping is that the ancillary can be priced closer to the maximum of the pricing range when the probability of purchase is high, and lower for low probabilities. Hence, a price mapping is chosen based on (1).

$$P^{rec} = \frac{L}{1 + \exp{}^{-}k(x - x_0)} \tag{1}$$

According to (1), three parameters can be controlled to map the price desirably.

- Max value, L : this is the full price of the ancillary
- Shape factor, k : the shape or steepness of the curve
- Mid point, x_0 : the mid point of the sigmoid curve

The shape factor k and mid-point x_0 can be fine-tuned to be either aggressive or conservative with pricing. This tuning is illustrated in Figure 6, indicating that at low purchase probabilities, the model compensates by reducing the recommended price.

Discrete Exhaustive Search

Exhaustive search can be efficiently performed over a small set of discrete prices that are within the pricing range. For a given probability of purchase $f_\theta(\boldsymbol{x}, P)$ and price P, expected revenue is computed using (2).

$$\hat{\mathbb{E}}_P = P \times f_\theta(\boldsymbol{x}, P) \tag{2}$$

Here f_θ is the learned function with θ as trainable parameters and $\boldsymbol{x}$ is the feature vector.

End-to-end Deep Neural Network

Suppose we are given N training samples $\{\boldsymbol{x}_i, y_i\}_{i=1}^{N}$, where $\boldsymbol{x}_i$ is the feature vector and y_i is the ground truth label for the i^{th} session. For purchased ancillaries, y_i equals 1 and 0 otherwise. The recommended price P^{rec} for feature vector $\boldsymbol{x}$ is denoted by $P^{rec} = \mathbb{F}_\Theta(\boldsymbol{x}, \mathbb{F})$, where Θ is a set of trainable parameters that can be learned for the mapping function $\mathbb{F}$, and $\mathbb{F}$ is a set of discrete price points in the pricing range.

The objective of the learning is to minimize the loss $\boldsymbol{\mathcal{L}}$ given as

$$\boldsymbol{\mathcal{L}} = \arg\min_{\theta} \sum_{i=1}^{N} \sum_{j=1}^{|\mathbb{F}|} (\Phi_{lb} + \Phi_{ub}) \cdot \mathbb{1}_{(\sigma_{ij} > 0)} \tag{3}$$

where the lower bound function Φ_{lb} and upper bound function Φ_{ub} are defined as,

$$\Phi_{lb} = \max\left(0, \left(L(P_{ij}, \delta_{ij}) - \mathbb{F}_\Theta(\boldsymbol{x}_i, \mathbb{F})\right)\right)$$
$$\Phi_{ub} = \max\left(0, \left(\mathbb{F}_\Theta(\boldsymbol{x}_i, \mathbb{F}) - U(P_{ij}, \delta_{ij})\right)\right)$$

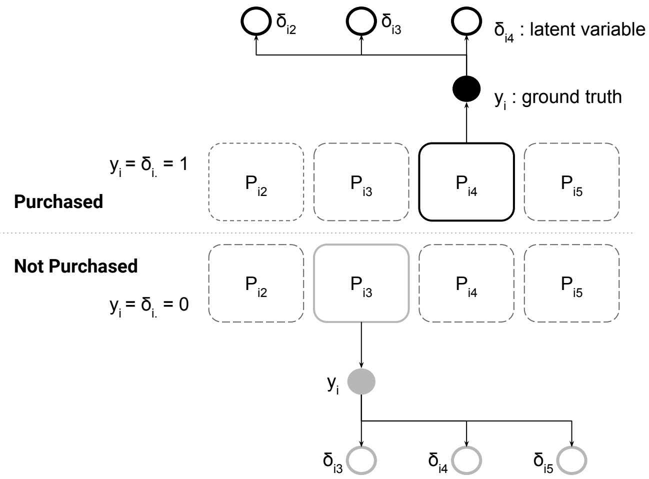

Figure 12 **Latent variable δ mapping from ground truth y**

where δ_{ij}, shown in Figure 12, is a latent variable that ensures the monotonicity in the willingness to pay assumption by taking the current ground truth y_i into account. The indicator function $\mathbb{1}_{(\sigma_{ij}>0)}$ selects loss values corresponding to those δ_{ij} which satisfy the monotonicity condition. Therefore, δ_{ij} is defined as

$$\delta_{ij}(y_i) = \begin{cases} y_i & \text{if } \sigma_{ij} \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{4}$$

Where, σ is the willingness to pay factor, defined as:

$$\sigma_{ij} = (j - j^*) \cdot (-1)^{y_i} \tag{5}$$

Assuming prices are listed in ascending order, j^* is the index at which P_{ij} equals $\mathbb{F}_\Theta(\boldsymbol{x}_i, \mathbb{F})$. We use L and U for the lower bound and the upper bound of the optimal price range, respectively. The functions $L(P_{ij}, \delta_{ij})$ and $U(P_{ij}, \delta_{ij})$ are defined as follows:

$$L(P_{ij}, \delta_{ij}) = \delta_{ij} \cdot P_{ij} + (1 - \delta_{ij}) \cdot c_1 P_{ij} \tag{6}$$

When the ancillary is purchased, the lower bound L is the purchase price P_{ij}. Otherwise, a lower price of $c_1 P_{ij}$ is set to be the lower bound, where $c_1 \in (0,1)$.

$$U(P_{ij}, \delta_{ij}) = (1 - \delta_{ij}) \cdot P_{ij} + \delta_{ij} \cdot c_2 P_{ij} \tag{7}$$

The upper bound U is P_{ij} when the ancillary is not purchased, whereas if the ancillary is purchased, a price of $c_2 P_{ij}$ $(c_2 > 1)$ is set as the upper bound.

Table 4 illustrates the lower and upper bound loss values for recommended price with respect to discrete price points. For $P_{ij} < \mathbb{F}_\Theta(\boldsymbol{x}_i, \mathbb{F})$, the upper bound loss increases linearly. For upper bound loss to be non-zero, $c_2 > \frac{\mathbb{F}_\Theta(\boldsymbol{x}_i, \mathbb{F})}{P_{ij}}$. Similarly, for non-zero loss, the bounds on c_1 are set to $\frac{\mathbb{F}_\Theta(\boldsymbol{x}_i, \mathbb{F})}{P_{ij}} < c_1 < 1$. For $c_1 = c_2 = 1$, the lower bound and upper bound are equal and hence the optimal price will be the j^{th} price in the price set $\mathbb{F}$. Therefore, c_1 and c_2 can be chosen to change the gap between the lower and upper bounds.

Table 4 Lower bound and Upper bound loss values

Prices	$\Phi_{lb} \cdot \mathbb{1}_{(\sigma_{ij}>0)}$	$\Phi_{ub} \cdot \mathbb{1}_{(\sigma_{ij}>0)}$
$P_{ij} < \mathbb{F}_\Theta$	0	$\max(0, \mathbb{F}_\Theta(\boldsymbol{x}_i, \mathbb{F}) - c_2 P_{ij})$
$P_{ij} = \mathbb{F}_\Theta$	0	0
$P_{ij} > \mathbb{F}_\Theta$	$\max(0, c_1 P_{ij} - \mathbb{F}_\Theta(\boldsymbol{x}_i, \mathbb{F}))$	0

Performance Metrics

Inspired from Ye et al. (2018), we leverage *Regret Score* (RS), *Price Decrease Recall* (PDR), and *Price Decrease Precision* (PDP) for comparing the pricing models as they are highly correlated with the airline's business metrics. PDR measures the likelihood of our recommended prices being lower than the current offered prices for non-purchased ancillary, and PDP measures the percentage of non-purchased ancillaries where the recommended price is lower than current offered prices. After our modeling phase, we have the original price $\mathbf{P}$ and the recommended price $\mathbf{P}_{\text{rec}}$ for every booking session. In addition, the actual ground truth is available i.e. whether the customer successfully purchased the ancillary in each session or not. Each possible case is listed in the table below (Table 5).

Table 5 Category of samples for metric computation

	Purchased	Not Purchased
$\mathbf{P}_{\text{rec}} \geq \mathbf{P}$	a	b
$\mathbf{P}_{\text{rec}} < \mathbf{P}$	c	d

Here, "a" denotes customer sessions where the customer purchased the ancillary but our model is recommending a higher price than shown previously, and "c" refers to those cases where the model is recommending a lower price than presented to the customer. On similar lines, "b" and "d" encapsulate those sessions where the customer didn't purchase the ancillary. We use these 4 values for computing the custom metrics mentioned previously. To illustration the metric calculation better, we randomly curate original and recommended prices for 8 sample sessions in Table 6. Booking status of 1 refers to a purchase, and 0 otherwise.

Table 6 Sample sessions with recommended prices

	S1	S2	S3	S4	S5	S6	S7	S8
$\mathbf{P}$	10	15	10	25	40	20	15	30
$\mathbf{P}_{\text{rec}}$	8	12	15	35	37	18	18	28
Booking Status	0	0	1	1	1	0	0	1

- **Price Decrease Recall (PDR)**: Among all sessions with no purchased ancillary, this metric calculates the percentage of recommendations that are lower than original prices.

$$\mathrm{PDR} = \frac{d}{b+d}$$

For the given example sessions, $d = 3$ (S1, S2, S6) and $b = 1$ (S7), so, PDR is **0.75**.

- **Price Decrease Precision (PDP)**: Among all sessions where $\mathbf{P}_{\mathrm{rec}} < \mathbf{P}$, PDP indicates the percentage of sessions that had no purchase.

$$\mathrm{PDP} = \frac{d}{c+d}$$

From the example in Table 6, $d = 3$ and $c = 2$ (S5, S8), so, PDP is **0.6**.

- **Regret Score (RS)**: Similar to booking regret in Ye et al. (2018), RS measures on an average how close our recommended price $\mathbf{P}_{\mathrm{rec}}$ was to the true purchase price $\mathbf{P}$.

$$\mathrm{RS} = \underset{purchases}{\mathrm{mean}}\left(\max\left(0, 1 - \frac{\mathbf{P}_{\mathrm{rec}}}{\mathbf{P}}\right)\right)$$

For the example, the purchases are S3, S4, S5, and S8. So, RS comes out to be **0.071**.

- **Price Decrease F1 (PDF1)**: This score is inspired by the F1 score used to evaluate the precision and recall trade-off in machine learnning. PDF1 therefore measures the trade-off between PDR and PDP.

$$\mathrm{PDF1} = \frac{2 \cdot \mathrm{PDR} \cdot \mathrm{PDP}}{\mathrm{PDR} + \mathrm{PDP}}$$

It follows from PDR and PDP scores previously : PDPF1 = **0.667**.

References

Allon G, Bassamboo A, Lariviere M (2011) Would the social planner let bags fly free? Working Paper.

Belobaba P, Odoni AR, Barnhart C (2015) *The Global Airline Industry* (John Wiley & Sons), 2nd edition.

Bockelie A, Belobaba P (2017) Incorporating ancillary services in airline passenger choice models. *Journal of Revenue and Pricing Management* 16(6):553–568.

Boeing (2013) Airline revenue management. Boeing conference presentation, `http://aviation.itu.edu.tr/\%5Cimg\%5Caviation\%5Cdatafiles/Lecture\%20Notes/FundamentalsofAirlineManagement20152016/Lecture\%20Notes/09-ITU_Revenue_Management.pdf`.

Chen MK (2016) Dynamic pricing in a labor market: Surge pricing and flexible work on the uber platform. *EC '16: Proceedings of the 2016 ACM Conference on Economics and Computation*, 455, URL `http://dx.doi.org/10.1145/2940716.2940798`.

Cui Y, Duenyas I, Sahin O (2016) Unbundling of ancillary service: How does price discrimination of main service matter?

Drucker H, Burges CJ, Kaufman L, Smola A, Vapnik V, et al. (1997) Support vector regression machines. *Advances in neural information processing systems* 9:155–161.

Gabaix X, Laibson D (2006) Shrouded attributes, consumer myopia, and information suppression in competitive markets. *Quarterly Journal of Economics* 121:505–540.

Garrow LA, Hotle S, Mumbower S (2012) Assessment of product debundling trends in the us airline industry: Customer service and public policy implications. *Transportation Research Part A: Policy and Practice* 46(2):255 – 268.

IdeaWorks (2015) Airline ancillary revenue projected to be 59.2 billion dollar worldwide in 2015. `https://www.ideaworkscompany.com/wp-content/uploads/2015/11/Press-Release-103-Global-Estimate.pdf`, ideaWorks Article.

Kahneman D, Tversky A (1979) Prospect theory: an analysis of decision under risk. *Econometrica* 47:263–292.

LeCun Y, Bengio Y, Hinton G (2015) Deep learning. *Nature* 521(7553):436–444.

O'Connell JF, Warnock-Smith D (2013) An investigation into traveler preferences and acceptance levels of airline ancillary revenues. *Journal of Air Transport Management* 33:12 – 21, ISSN 0969-6997, URL `http://dx.doi.org/https://doi.org/10.1016/j.jairtraman.2013.06.006`, papers in Honor of Christina Barbot.

Parker DB (1985) Learning logic. *Technical Report, MIT Press* .

Rish I, et al. (2001) An empirical study of the naive bayes classifier. *IJCAI 2001 workshop on empirical methods in artificial intelligence*, volume 3, 41–46.

Shukla N, Kolbeinsson A, Otwell K, Marla L, Yellepeddi K (2019) Dynamic pricing for airline ancillaries with customer context. *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, KDD 2019, Anchorage, AK, USA, August 4-8, 2019*, 2174–2182 (ACM), URL `https://doi.org/10.1145/3292500.3330746`.

Shulman J, Geng X (2013) *Management Science* 59:899–917.

Taneja N (2016) *Designing Future-Oriented Airline Businesses* (Taylor & Francis), ISBN 9781317152163, URL `https://books.google.com/books?id=VEAHDAAAQBAJ`.

Tuzovic S, Simpson MC, Kuppelwieser VG, Finsterwalder J (2014) From 'free' to fee: Acceptability of airline ancillary fees and the effects on customer behavior. *Journal of Retailing and Consumer Services* 21(2):98 – 107.

van Ryzin GJ, Talluri KT (2005) *An Introduction to Revenue Management*, chapter 6, 142–194. URL `http://dx.doi.org/10.1287/educ.1053.0019`.

Vermeulen AH, Bezos JP (2010) Method and system for dynamic pricing of web services utilization. US Patent 7,743,001.

Werbos P (1974) Beyond regression: new tools for prediction and analysis in the behavioral sciences. *Ph.D. Dissertation, Harvard University* .

Ye P, Qian J, Chen J, Wu Ch, Zhou Y, De Mars S, Yang F, Zhang L (2018) Customized regression model for airbnb dynamic pricing. *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, 932–940 (ACM).