



## Predicting Customer Booking Cancellation Using Machine Learning

Data science use case  
for Travel and Hospitality industry

## Table of Contents

1. Introduction
2. ACS Solutions Approach
3. Data and Insights
4. Methods
5. Future Scope

## Executive Summary

Travel companies have to plan ahead to optimize their expenses. They utilize the significant manpower and time to execute the trips. The better understanding of the customer behavior would help them to avoid the last minute cancellations and also to reduce the costs. They maintain customer's personal and trip information and their browsing activity information.

The White Paper covers how Machine Learning helps Travel companies to identify the customers who are likely to cancel their trips after registration.

### Lead Authors

Chandranth Vemula [Data Scientist](#)

### Contributors

Vamsi Krishna Nori [Data Scientist](#)

Manisha Kallepalle [Data Scientist](#)

Vivek K [Data Scientist](#)

### Guided by

Manas Pant [Manager](#)

Venkat Palla [Sr. Manager](#)

### About ACS Solutions

ACS Solutions is a leading global information technology services and consulting organization that has been serving businesses globally across industries since 1998. A trusted partner to both mid-market and Fortune 500 clients, ACS Solutions has been instrumental in each of their unique digital transformation journeys. Our extensive industry-specific domain expertise and passion for innovation has helped clients envision, build, scale and run their businesses more efficiently, for over two decades.

We have a proven track record of developing large and complex software and technology solutions for Fortune 500 clients such as Microsoft, Blue Cross Blue Shield, Cox Communications, and Novartis. We deliver on our commitments and enable our customers to achieve a digital competitive advantage through flexible and global delivery models, agile methodologies, and expert frameworks. Headquartered in Duluth, GA, and with several locations across North America, Europe, and the Asia-Pacific regions, ACS Solutions specializes in 360-degree digital transformation and IT consulting services.

For more information, please reach us at – [acssolutions@acscorp.com](mailto:acssolutions@acscorp.com)

Copyright © 2018 [ACS Solutions](#)  
All rights reserved.

## Introduction

The tourism industry has expanded a lot in the recent years and is continuing to do so as reported in the WTTC's (World Travel and Tourism Council) latest annual research that "Travel and Tourism's contribution to world GDP outpaced the global economy for the sixth consecutive year in 2016, rising to a total of 10.2% of world GDP (US\$7.6 trillion). The sector now supports 292 million people in employment – that's 1 in 10 jobs on the planet."

Travelers prefer to make a significant part of their travel arrangements in advance. These include flight bookings, hotel reservations, local conveyances, activities to do, local places to check out. They want to contact local travel guides, check reviews and compare prices of different vendors at one place instead of going to multiple sites. Online Travel Agencies, also known as OTAs, provide this platform. They also enable travelers to rank different packages by facilities offered and proximity to a preferred city or landmark. Handling this vast business is not an easy task as OTAs face many challenges. They need to maintain price parity and booking synchronization with different hotels and airline companies. They must get in live feed from different vendors when a customer asks for availability and also update the vendor for every booking and cancellation done. Moreover, different hotels and airlines have different cancellation policies. Many times, customers have unforeseen changes in their plans and then they go for 'Cancellations'. OTAs suffer a considerable amount of loss because of bookings getting canceled.

Online Travel Agencies, also known as OTAs, provide platform to contact local travel guides, check reviews and compare prices of different vendors at one place instead of going to multiple sites.



Sometimes customers cancel their booking only a few days prior to the trip. At that point of time, an OTA has already made many arrangements for the customer like allocating resources, making a booking at the local restaurants and sites that those customers were supposed to visit during the trip, arranging places for their stay and paying in advance to local guides. Consequently, all those efforts and advance arrangements go to waste.

**But what if we can predict which reservations are likely to get canceled ?**

**By predicting the probable cancellations, OTAs can:**

**Increase Reservation Retention:** Know which reservations are going to get canceled which helps them to prevent cancellations by giving those specific customers some extra benefits or discounts.

**Estimate better net demand forecast:** Eliminate the reservations which are likely to be canceled which enables OTAs to have a better actual estimate of demand which further helps them in price modeling and resource allocation.

**Better Cancellation and Offer policies:** Understand the variables correlated with cancellation so that the OTA can decide offer strategies and cancellations charges accordingly.

**By predicting the probable cancellations, OTAs can**



**Increase Reservation Retention**



**Estimate better net demand forecast**



**Better Cancellation and Offer policies**



## How do OTAs use Data Science already to make crucial business decisions?

OTAs have already been using data analytics tools for tasks like customer segmentation, customer profitability and revenue estimation. With the advancement of Machine Learning and Artificial Intelligence, they began to incorporate these modern technologies with their existing analytics architectures.

They now do, not just descriptive analytics, but also predictive analytics like Customer lifetime value estimation and Customer churn prediction. Moreover, they have not limited themselves to just Customer analytics. They are utilizing ML for many purposes like price modeling of flights and trips and recommending customized packages to their customers.

A few have started using NLP techniques to create chat-bots who act as travel agents and answer customer queries in a much more effective way. OTAs are also trying to understand their customers by having intelligent algorithms to go through the whole databases of past customers searches, trips taken etc.

Revenue management, optimized disruption management and understanding customer feedback through sentiment analysis of customer reviews on websites and social media are some of the other applications of ML in this domain.

These technologies have made planning and booking process a lot easier as customers are finding relevant trips according to their needs and chat-bots are helping to reduce the turnaround time for customer queries

## ACS Solutions Approach

ACS Solutions has been working in collaboration with travel companies for many years now and has gained a valuable experience in this domain. We find insights about their customer preferences so as to suggest them relevant trips and packages through emails. This has helped our partners to increase their customer satisfaction and loyalty. Personalized content emails and reminders encourage customers to return, reduce churn and increase their lifetime value.

In this project, we partnered with an existing client and applied ML to predict the reservations which are going to get canceled and after how many days of booking those reservations will get canceled.

We started off with basic approaches but simple methods did not perform well with this data-set. This is not much of a surprise as we observed that there is a class imbalance in the data, meaning the proportions of reservations which get canceled is low. More robust approaches like Ensemble and Neural Nets work well even with class-imbalanced data.

We got significant accuracy in prediction on this problem using two very popular approaches, namely Gradient Boosting and Neural Networks.



## Data and Insights

We had their reservation data, customer's data and data of their customer's interaction with the company either online or through phone calls. Reservation data is comprised of information like reservation date-time, trip details etc. Customer data includes every customer's personal information like age, gender, demographics etc. and their previous travel information like the number of trips taken by them in the past, what kinds of trips they booked in the past etc. We took these kinds of data and after exploratory analysis, we found out many insights. Some of them are:

- We observed that around 20% of reservations were canceled every year.
- A reservation made well in advance has higher chances of cancellation.
- If an insurance is taken on a reservation then it has less chances of getting canceled.
- If a customer has canceled more number of reservations in the past then he is more likely cancel his current reservation again.
- Reservations made by first-time customers have higher cancellation chances.

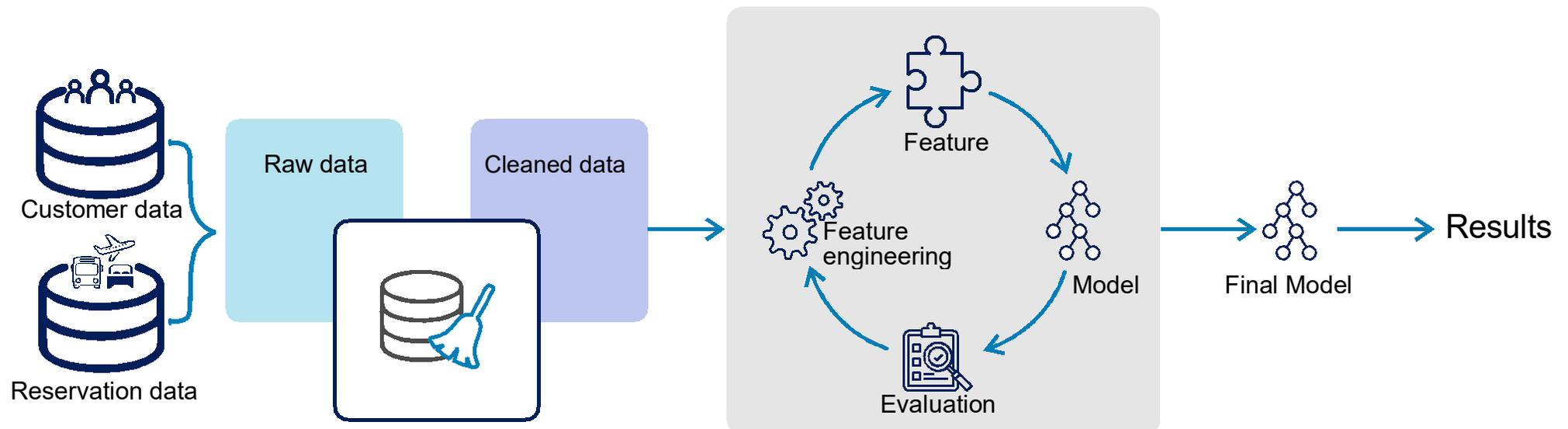
Machine learning algorithms use these correlations between outcome (reservation final status) and attributes (reservation information) to learn patterns and make predictions. For example, for a reservation, the algorithm looks at the attributes like advance booking days, insurance flag etc. and decide the outcome for that reservation, it also tells us the attributes which were most helpful for it to decide the outcome and ranks them according to their relative importance. The model deployed by us is not specific to the dataset that we have used. It is an accelerator which can be easily tweaked a little to work with different datasets so that other companies can also utilize it.

## Methods

### Gradient Boosting

Gradient boosting is a very popular technique among people who do machine learning on a regular basis. It is an ensemble method where instead of training a single model, we train multiple independent base models. What makes boosting more interesting and different from other ensemble methods is that here every new model learns from

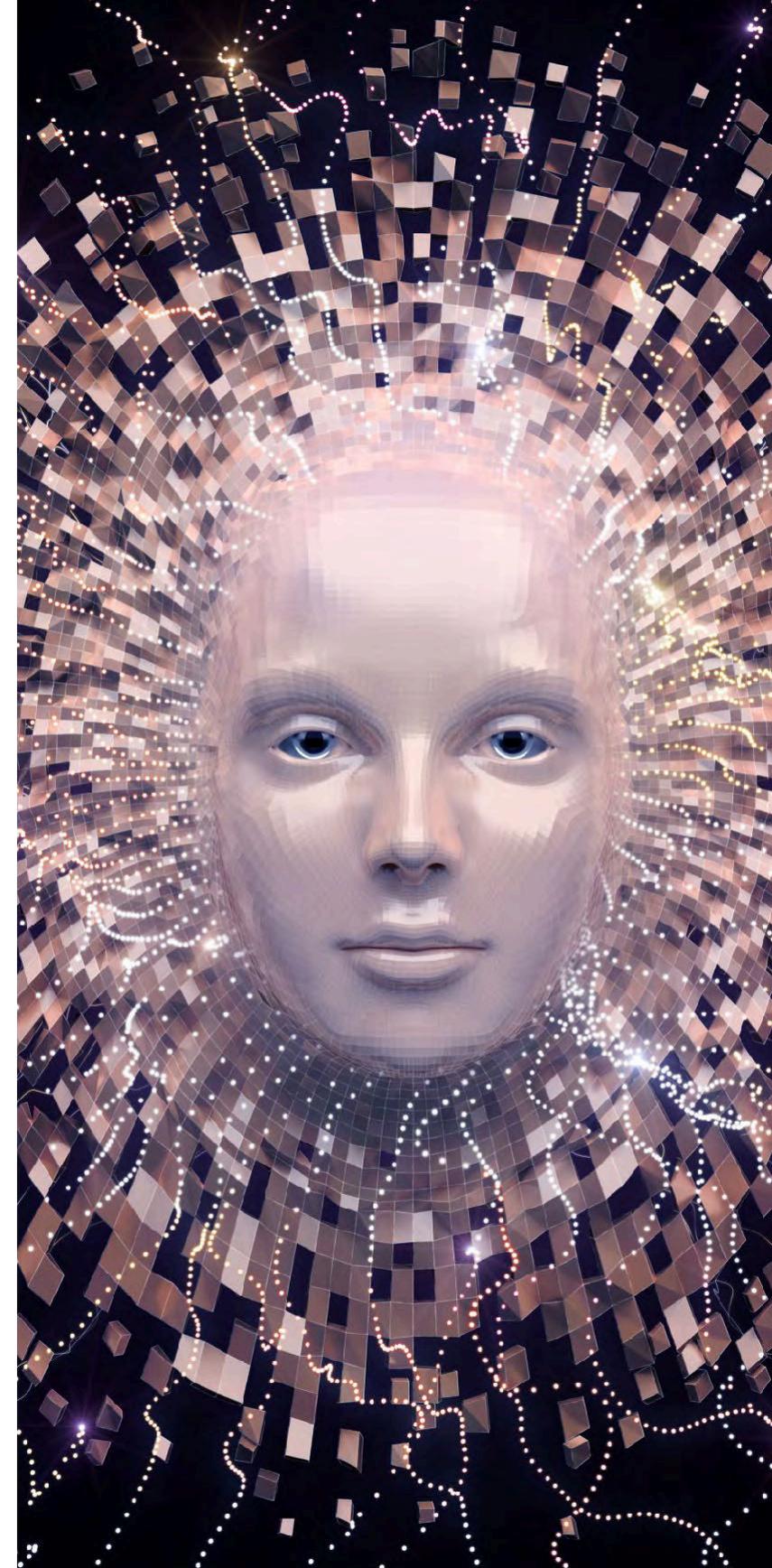
the mistakes done by its previous model. We took binary decision trees as our base models for boosting. A binary decision tree learns to split data into subsets at every node such that the subsets are more homogenous in terms of the target class. Many such trees are trained in a sequential manner and are used for final prediction.



## Neural Networks

It is a biologically inspired model. It is a network of neurons like those in our body. A Neural Network model takes all combinations of inputs to create new layers and it learns from the training data to figure out the weight of connection between the nodes. It considers the importance of the right combination of attributes.

Every node gets its value from a linear combination of value of nodes in the previous layer multiplied by their respective weights. It reaches the outcome in the same sequential manner. If the outcome is inaccurate, it goes back to change the weights in a manner such that when it comes forward in the next iteration it makes a more accurate prediction than the last one.



## Future Scope

These, however, are just the beginning steps. Prediction of trips vulnerable to cancellation can also be used to filter out list of actual users versus the 3rd party users who make bulk reservations well in advance to sell it to other customers during peak time with increased price and will cancel the bookings for which there is no buyer.

Implementation of a more advanced chatbots system which makes smart guesses of what a customer is going to ask next can speed up the process of customer service execution even further.

Alerting and monitoring is something which can be eased using Machine Learning. Travel Industry generates a huge volume of data per day which could not be processed by human beings. Using highly supervised machine learning algorithms, by observing different sets of factors, problems can be anticipated very like how a doctor diagnose a disease by knowing a set of symptoms.

**Fraud detection:** More than 3 billion passengers travel by air every year across the world resulting in generation of huge online transactions. There are many high-value frauds in payments and transactions, leading to a significant loss. Hence, there is a need to validate these transactions in real time.

