Forecasting Air France KLM passenger traffic

Improving passenger traffic prediction algorithm accuracy for a better operational effectiveness.

M. Tech Thesis

For the certification of Data Science study at Telecom ParisTech

Mame-Mor FALL

December 19,2016

Telecom ParisTech - Mines-Telecom institute

Preface

I started this project in August 2016 for Telecom Paris Data Science certification. This project has been a great learning experience in both a personal and professional perspective.

The topic of this project is a pure machine learning subject and exactly what I was looking for in the frame of my study. The challenge of beating the Department of Operational Research current algorithm on passenger traffic prediction has been thrilling. I hardly imagine a more strategic subject with huge impact throughout our group and which stakes are easily understandable or applicable to any company.

I would like to thank Isabel Gomez, head of Air France KLM Department of Operational Research, for introducing me to her team and I would like to express my gratitude to Blaise-Raphael Brigaud and Wiem Elgazhel who supported me all along the project with their expert knowledge, their guidance and their encouragement.

Mame-Mor Fall

Table of Contents

1 - Introduction

1-1 Research context

Air France - KLM is a world leader in passenger transportation and cargo transportation with more than 2200 daily flights enabling its customers to travel to 320 destinations in 114 countries. In 2015 the group carried a total of 89.8 million passengers and 1.2 million tons of cargo on board its 534 aircrafts. Within the group exists several airlines :

- Air France,
- KLM,
- HOP*!* for short-haul flights from and to France,
- KLM Cityhopper its counterpart from Netherland,
- Transavia France and Netherland the low cost carriers of the group.

And the Air France - KLM group (we will use the abbreviation AFKL from now on) is itself within a global airline alliance called Skyteam with Delta Airlines (largest american airline), Alitalia, Aeroflot (largest russian airline), China Southern/Eastern etc.
For all these companies, one of the most essential activity in our operations is the planning part. For passenger traffic we have to plan :

- the number of staff needed in our main station to assist the passenger 4-5 weeks before a flight (this is true for AFKL flights but also for our partners for which we handle the flights in our stations),
- the number of staff needed to handle the luggages 3 weeks before,
- the number of meals per class needed inside the aircraft for our passengers 1 week before a flight (catering),

The majority of our cargo traffic uses the space left inside the bellies of the passenger flights, so planning the number of passenger luggages is critical also for cargo traffic.

To make these decisions AFKL relies on its Operation Research Department (OR) within the IT Division. The model used for these predictions is called PTRA for Passenger TRAffic. It has been built 10 years ago and is considered probably outdated in comparison to modern computer development and machine learning theory breakthroughs by the OR management. On top of this the operational executives using the model are challenging for new developments to improve prediction accuracy which could potentially results in significant spending savings. The Operation Research Department wants to validate their hypothesis that the algorithm is beatable before spending time and resources to overhaul the prediction system.

1-2 Scope of this research

The primary aim of my research is to prove that PTRA is beatable on passenger traffic prediction and passenger traffic in transfer prediction. The differences between the two is

that passenger traffic include all traffic when passenger traffic in transfer means a passenger which has a continuing flight after the AFKL operated flight.
For that I have been given all data on 2 Origin-Destination pairs:

- Flights between Paris CDG to New-York JFK
- Flights between Paris CDG and Lisbon LIS

We will detail these data later.
The meaning of “beatable” needs further clarification as the operational cost function related to PTRA precision error is complex. This will be discussed later, for now let’s just consider “beatable” as being generally more accurate consistently than PTRA algorithm.

The second aim of my research is to lay the foundation of a scalable code which will be put in production with the help of another intern coming in April 2017 within the OR Department.

1-3 Structure of this document

The document starts by a summary of traffic prediction literature review in chapter 2. Which has been the very start of my research project. In Chapter 3 the project plan outlines in details the context in which the project has taken place and the fundamental questions we will try to answer. Chapter 4 to 8 presents the different steps followed within the frame of the project to understand the problems, build data models, visualize results and validate my hypothesis or reiterate this process to newly found problems. Chapter 9 makes a detailed case for the final results of this research project and chapter 10 emphasize the findings, conclusions and recommendations for the project continuation.