

Predicción y Análisis del TOBT

Grado en Gestión Aeronáutica
Proyecto de Fin de Grado
realizado por
Aitor Serrano Rico
y dirigido por
Liana Napalkova
Sabadell, 09 de Febrero de 2017

HOJA DE RESUMEN – TRABAJO DE FINAL DE GRADO. ESCUELA DE INGENIERIA

<i>Análisis y Predicción del TOBT</i>	
<i>Predicció i Anàlisi del TOBT</i>	
<i>TOBT Analysis and Prediction</i>	
Autor[a]: <i>Aitor Serrano Rico</i>	Data: <i>Febrero del 2017</i>
Tutor[a]/s[es]: <i>Liana Napalkova</i>	
Titulació: <i>Grado en Gestión Aeronáutica</i>	
Palabras clave	
<ul style="list-style-type: none">• Catalán: <i>Retards, TOBT, ACDM, predicció, anàlisi</i>• Castellano: <i>Retrasos, TOBT, ACDM, predicción, análisis</i>• Inglés: <i>Delays, TOBT, ACDM, prediction, analysis.</i>	
Resumen del Trabajo Final de Grado	
<ul style="list-style-type: none">• Català: El projecte realitzat, tracta d'un estudi dels temps de les aeronaus durant el seu procés de trànsit en l'aeroport, amb l'objectiu de conèixer els motius dels retards del 'TOBT'. Per arribar-hi, es revisen tots els processos que poden influir en els 'turnarounds' dels avions, provocant retards en la sortida del vol i també l'impacte, en tots els aspectes, d'aquests retards. Tot això, amb la finalitat de crear un model predictiu real, utilitzant dades reals de tres companyies aèries, Americanes, diferents (American Airlines, Alaska Airlines i Aloha Airlines).• Castellano: El proyecto realizado se trata de un estudio de los tiempos de las aeronaves durante su proceso de tránsito en el aeropuerto, con el fin de conocer los motivos de los retrasos del 'TOBT'. Para ello, se revisan todos los procesos que pueden influir en el 'turnarounds' de los aviones, provocando retrasos en la salida del vuelo y también del impacto, en todos los aspectos, de estos retrasos. Todo ello, con el objetivo final de crear un modelo predictivo real, utilizando datos reales de tres aerolíneas, Americanas, diferentes (American Airlines, Alaska Airlines y Aloha Airlines).• English: The given thesis is aimed at the study of the time that an aircraft spend during the transit at the airport, in order to know the reasons for the delays of 'TOBT'. To do this, the thesis will review all processes that can influence the turnaround of the aircraft, causing delays in the flight departure and also the impact, in all aspects, of these delays. All this, with the aim of creating a real predictive model, using real data from three different American airlines (American Airlines, Alaska Airlines and Aloha Airlines).	

TABLA DE CONTENIDOS

INTRODUCCIÓN.....	9
Estado del arte y motivación.....	9
Objetivos y tareas del proyecto.....	10
Valor Práctico.....	10
Metodología.....	11
Estructura del Proyecto.....	11
1. DEFINICIÓN DEL PROBLEMA: ESTUDIO Y ANÁLISIS DEL ‘GROUND HANDLING’	12
1.1.- Definción de ‘Groud Handling’	12
1.2.- Retrasos de ‘Handling’	17
1.3.- Definición del ‘Turnaround’ de las Aeronaves.....	19
1.4.- Retrasos en el ‘Turnaround’	22
1.5.- El proceso ‘Turnaround’	25
1.5.1.- Antes de la llegada.....	25
1.5.2.- ‘On the ground’	28
2.- ACDM: AIRPORT COLLABORATIVE DECISION MAKING.....	34
2.1.- ACDM.....	34
2.2.- ¿Por qué ACDM?	35
2.3.- Beneficios del ACDM.....	36
2.4.- Beneficios de Predictibilidad en la Llegada.	37
2.5.- Principales elementos del A-CDM.....	40
2.5.1.- Procedimiento TSAT.....	40
2.5.2.- TOBT UPDATE.....	42
3. ANALISIS DE LOS MODELOS PREDICTIVOS.....	45
3.1.- Modelo predictivo ‘Random Forest’.....	47
3.2.- Modelo predictivo ‘Gradient Boosting Machine’.....	49
3.3.- Evaluación del modelo usando RMSE (Root-Mean-Square Error).....	52

4.- CASO DE ESTUDIO: PREDICCIÓN DEL TOBT.....	53
4.1.- Proceso de limpieza de datos.....	56
4.2.- Pre procesamiento de datos.....	59
4.2.1.- Asimetría (Skewness).....	62
4.2.2.- Centrado y Escalado de las variables.....	64
4.3.- Selección de las características.....	65
4.4.- Creación del Modelo predictivo.....	69
4.4.1.- Gradient boosting.....	70
4.4.2.- 'Random Forest'	71
CONCLUSIONES.....	73
BIBLIOGRAFIA	75
ANEXO 1: DATA_CLEAN.IPYNB.....	77
ANEXO 2: PREPROCESING.IPYNB.....	78
ANEXO 3: PREDIVTIVE_MODEL.IPYNB.....	81

TABLA DE ILUSTRACIONES

- Figura 1 - Costes aproximados por minuto de retraso*
- Figura 2 - Proceso secuencial del 'Turnaround'*
- Figura 3 - Diagrama de actividades del 'Turnaround'*
- Figura 4 - Cadena de valor de Porter*
- Figura 5 - Representación gráfica del 'Turnaround'*
- Figura 6 - Representación de la colocación de los equipos de 'Handling'*
- Figura 7 - Representación llegada en hora y salida en hora*
- Figura 8 - Representación llegada en hora y salida en hora (Buffer time)*
- Figura 9 - Representación llegada en hora y salida adelantada*
- Figura 10 - Representación llegada tardía y salida en hora (Buffer time)*
- Figura 11 - Representación llegada tardía y salida en hora (Durante Buffer time)*
- Figura 12 - Representación llegada tardía y salida tardía*
- Figura 13 - Representación llegada en hora y salida tardía*
- Figura 14 - Planificación (en diagrama de barras) de tiempos de 'Turnaround'*
- Figura 15 - Ejemplo. Revisión de SLOT*
- Figura 16 - Error en la descarga, al no respetar el balance de la aeronave.*
- Figura 17 - Ejemplo de planificación en diagrama para embarque en parking remoto*
- Figura 18 - Ejemplo hoja de carga. Boeing 757*
- Figura 19 - Información compartida por los actores del ACDM*
- Figura 20 - Mecanismos de beneficio de A-CDM y ejemplos de facilitadores*
- Figura 21 - Estabilidad de la utilización de stands en el aeropuerto de Frankfurt en 2015*
- Figura 22 - Impacto del A-CDM en la previsibilidad del tiempo de llegada*
- Figura 23 - Factores que afectan al valor TSAT y ejemplos de métodos de publicación*
- Figura 24 - Tendencia de cumplimiento de TSAT en Londres Heathrow desde octubre 2012 hasta el 2015.*
- Figura 25 - Representación visual de participantes en el TOBT*
- Figura 26 - Fases del modelado predictivo*
- Figura 27 - Árbol de Random Forests*
- Figura 28 - 'Flow chart' del Random Forest*
- Figura 29 - Gradient Boosting Machine*
- Figura 30 - Metodología GBM*

Figura 31 - Demostración cómo la complejidad del modelo influye tanto en la precisión de la predicción y el rendimiento computacional.

Figura 32 - Impresión de variables que contienen alguna muestra 'N/A'

Figura 33 - Comprobación variables con 'N/A'

Figura 34 - Cabecera base de datos (observaciones con valores '0')

Figura 35 - Cabecera base de datos. Limpieza de datos.

Figura 36 - Grafico de barras. Horas de salida.

Figura 37 - Grafico de barras. Horas de llegada.

Figura 38 - Grafico de barras. Mes del vuelo.

Figura 39 - Grafico de barras. Día de la semana.

Figura 40 - Grafico de barras. Tiempo de vuelo.

Figura 41 - Grafico de barras. Distancia de vuelo.

Figura 42 - 'SKEWNESS'. Ejemplo grafico sin transformar

Figura 43 - 'SKEWNESS'. Ejemplo grafico transformado

Figura 44 - Output: Media + Desviación STD.

Figura 45 - Output: Media + Desviación STD (Transformada).

Figura 46 - Importancia relativa de las variables. Gráfico.

Figura 47 - Importancia relativa de las variables. Valores.

Figura 48 - Proceso entrenamiento 'Gradient Boosting Model'

INTRODUCCIÓN

Estado del arte y motivación

La creciente actividad de la industria aérea ha estado acompañada de un empeoramiento creciente de la puntualidad de los vuelos y con ello de la calidad del servicio que se presta, y por eso se trata de una necesidad analizar los diversos enfoques frente a esta realidad. Se han investigado extensamente las causas que existen tras el aumento creciente de la congestión en los aeropuertos y los atrasos y cancelaciones de vuelos. Dentro de las distintas líneas de investigaciones se encuentran argumentos tales como las externalidades de congestión en los aeropuertos.

En la actualidad existe una creciente saturación en los aeropuertos principalmente en las horas de mayor frecuencia así como en fechas específicas, donde se hace indispensable la necesidad del control de los tiempos del 'turnaround', de ello dependerá que las entradas y salidas de las aeronaves sucedan con fluidez, evitando así retrasos y otros inconvenientes derivados de estos retrasos. Para ello es importante observar y determinar qué factores toman un papel importante en el tiempo que conlleva el cambio de vuelo, para ello, no se puede dejar a la intuición, ni a la vez a estudios poco rigurosos, y se deberá establecer procesos estadísticos que determinen con el menor error posible cuáles son esas variables y cómo se puede reducir el impacto en el tiempo del 'turnaround'.

A modo de control frente a este constante crecimiento del movimiento en este sector, es totalmente necesario el tener un seguimiento continuo sobre las herramientas de mejora de cualquier proceso que pueda influir en los tiempos de tránsito de las aeronaves en el aeropuerto.

El mercado aeronáutico se está encaminando principalmente por las soluciones 'Low Cost', estas compañías operan con un modelo de tiempos de escala, mucho más estrechos y sin margen que una compañía de bandera por ejemplo. Esto es debido a que cada minuto que el avión no está en el cielo, es dinero que pierde y/o deja de ganar la compañía, es por ello que es uno de los principales objetivos que persiguen las aerolíneas que venden los billetes a precios tan económicos, esos precios no serían posibles si se ocasionan retrasos que no permitan que el avión tome vuelo con la mayor celeridad.

Los retrasos que sufren los vuelos, en la práctica están generados por tres causas principales: la capacidad aeroportuaria y su utilización, los esfuerzos que realizan las aerolíneas en puntualidad, y las interacciones entre vuelos. Así que una parte residirá en los esfuerzos que haga otro agente, pero hay que recordar que la otra parte a la vez tendrá el interés de reducir el tiempo en el 'turnaround'.

La predicción de los tiempos de 'turnaround' es de gran importancia para las compañías aéreas, para que se lleve a cabo con éxito se necesitarán cada vez inputs de mayor calidad para que los resultados puedan ser usados con total certeza en la toma de decisiones de cualquier aerolínea, pudiéndose usar como herramienta, las herramientas que hoy en día son usadas en la Ciencia de Datos, y que evolucionan forma vertiginosa, obteniendo con ellas factores que no son visibles, así como correlaciones entre variables, entre otros.

Objetivos y tareas del proyecto

El objetivo principal del estudio es la identificación de los factores que impactan con mayor fuerza en los tiempos de 'turnaround', así como una gestión automatizada de la obtención de datos relacionados con estos tiempos, a través de aplicativos que distribuyen informaciones relativas a los tiempos de los vuelos.

Por otro lado será importante determinar los métodos estadísticos que se llevaran a cabo para el correcto procesamiento de la información obtenida. Para ello, se realizará un estudio sobre una base de datos extraída de datos reales de tres compañías americanas (American Airlines, Alaska Airlines i Aloha Airlines), acondicionándola para su mayor eficacia en el estudio del modelo predictivo.

Por último se deberá establecer qué modelo predictivo ofrece los resultados más precisos y con menor cantidad de error.

Para llevar a cabo el estudio, una visión general de las tareas a realizar serán las siguientes:

- Analizar los diferentes aspectos del 'ground handling' y los factores en los que interfiere
- Identificar el problema principal
- Seleccionar los entornos y los lenguajes con los que se trabajara para la generación de los códigos.
- Elaborar el procedimiento de obtención de datos, con su posterior limpieza y depuración de datos
- Identificar qué atributos, variables, o características son las de mayor relevancia, cuales afectan más a los tiempos del 'turnaround', y a la vez, cuáles pueden ser prescindidas del estudio ya que no tienen suficiente impacto en estos tiempos.
- Aplicar métodos estadísticos que permitirán la buena ejecución de un modelo predictivos
- Generar los resultados del modelo predictivo.

Valor Práctico

El valor de este proyecto es la intención de mejorar la operativa aeronáutica general mediante el correcto análisis, predicción y optimización de los procesos. Con ello, se puede reducir los costes tangibles e intangibles provocados por los retrasos. La implementación de dichos análisis y herramientas de predicción en la operativa aeroportuaria, permitirán conocer a fondo en qué manera afectan los factores y las variables al correcto desarrollo de cualquier acción dentro de la aviación y conocer sus consecuencias, así como crear predicciones altamente precisas de cuándo los vuelos van a retrasarse, cancelarse o demorarse, aportando un alto valor práctico.

Las herramientas que permitirán de forma automatizada, el análisis en la operativa aeroportuaria del momento, dando a conocer en profundidad de qué forma afectan los factores y las variables al correcto funcionamiento de las partes implicadas en el proceso del 'turnaround'. Esto, ante la creciente congestión que sufren los aeropuertos, está enfocado para la reducción de costes, a la vez que aprovechar de la mayor forma posible a la ventaja inherente que tiene la aviación, que es la velocidad.

Metodología

Para la elaboración del proyecto se han utilizado una serie de herramientas y procesos con tal de lograr el máximo rendimiento de éste y poder obtener unos resultados óptimos.

En primer lugar se ha realizado un análisis y minería de datos de tres aerolíneas americanas, seleccionando aquellas variables importantes para la implementación, así como los diferentes métodos de procesamiento e ingeniería de datos para limpiar y mejorar la eficiencia y efectividad, como el suavizado de datos y la detección de anomalías, con el fin de mejorar la calidad de los datos a utilizar.

Para el almacenamiento de los datos se usa el formato de archivo '.CSV' (Comma Separated Values), donde cualquier programa con hoja de Cálculo, es capaz de abrirlo y trabajar con la información que contiene.

Tanto el modelo predictivo como el procesamiento y acondicionamiento de datos se ha desarrollado mediante el programa 'Phyton', con su aplicación 'Jupyter Notebook'. 'Phyton' es un lenguaje de programación interpretado cuya filosofía hace hincapié en una sintaxis que favorezca un código legible. Se trata de un lenguaje de programación multi-paradigma, ya que soporta orientación a objetos, programación imperativa y, en menor medida, programación funcional. Esto permitía, a la vez que se hacía el procesamiento de datos, graficar datos y hacer estudios visuales mediante estudios de gráficos de los datos de las aerolíneas.

El pre-procesamiento de datos, se ha realizado mediante código 'Phyton', leyendo este el fichero '.csv' y generando más tarde otro fichero (también '.csv') con los datos modificados nuevamente, siendo más eficaces para el proyecto. Para la siguiente etapa del proyecto, se cogería este nuevo fichero generado y así sucesivamente, siendo el último fichero creado/modificado el más óptimo para el modelo predictivo.

Para concluir la parte de programación se usaran dos modelos predictivos, 'Random Forests' y 'Gradient Boosting Machine' para determinar, cual ofrece una predicción más precisa y con el menor índice de errores.

Una vez realizado esto, se extraerán las conclusiones sobre los datos ofrecido por los modelos de predicción.

Estructura del Proyecto

- **Introducción:** Descripción de la necesidad de la predicción de tiempos y retrasos
- **La definición del problema:** Análisis de la situación actual e identificación de carencias
- Estudio y Análisis del A-CDM y su importancia para el proyecto
- **Estudio del Caso:** Elaboración del estudio y modelo predictivo
- **Conclusiones**
- **Bibliografía**
- **Anexo:** Códigos utilizados en el modelo predictivo

1. DEFINICIÓN DEL PROBLEMA: ESTUDIO Y ANÁLISIS DEL 'GROUND HANDLING'

El retraso en el proceso de 'turnaround' está influenciado por la cantidad de tiempo reservada para este, la puntualidad en la llegada, entre otros, así como la eficiencia operacional de los servicios en tierra. Según la conclusión de la 'Performance Review Unit' (PRU), indica que principalmente las llegadas con retrasos son originadas principalmente por salidas con retrasos. Esto conduce a la propagación de retrasos en toda la rotación de aeronaves y a la vez en toda la red de una compañía aérea, un retraso, provoca un nuevo retraso.

Según Eurocontrol el coste medio por minuto para la aerolínea por un retraso es tanto cuando el avión de pasajeros está retrasado en suelo o en el aire, y es dividido en dos valores. Estos dos valores se diferencian por el hecho de que un valor es el coste directo para la aerolínea, y el restante es resultante de un coste indirecto teniendo en cuenta por ejemplo el coste de oportunidad de los pasajeros. [1]

Direct cost to an airline (for system wide use or individual airline studies)

	Delay cost per minute (€)	
Tactical with network effect (includes reactionary delay)	Ground	49.5
	Airborne	69.2
Strategic (buffer built into schedule)	Ground	16.7
	Airborne	51.0

(Adjusted to 2014 values)

Overall cost to an airline (for individual airline studies, includes passenger opportunity cost)

	Delay cost per minute (€)	
Tactical with network effect (includes reactionary delay)	Ground	90.8
	Airborne	110.5

(Adjusted to 2014 values)

Figura 1 - Costes aproximados por minuto de retraso

1.1.- Definición de 'Groud Handling'

El término 'Ground Handling' se refiere a una serie compleja de los procesos que se requieren para separar las fases de un avión desde su llegada a un aeropuerto y su salida. [5]

En el pasado, las aerolíneas solían realizar ellas mismas los servicios de personal y equipo. Con la desregulación y la necesidad de ser más conscientes de los costes, las aerolíneas comenzaron a "agrupar" aquellas actividades que podrían ser realizadas más baratas por empresas especializadas, externas. Esto ha llevado a la externalización de actividades de asistencia en tierra a empresas especializadas, a otra compañía aérea que goza de una ventaja comparativa en el manejo o en un aeropuerto. Estos servicios de asistencia en tierra están enumerados en el Manual de Manejo de Aeropuertos de la Asociación Internacional de Transporte Aéreo (IATA) y constituyen once categorías principales de actividades, junto con sus subcategorías, de la siguiente manera:

‘Handling’ y supervisión en tierra

- Servicios de representación y enlace con las autoridades locales o cualquier otra entidad, desembolsos en nombre del usuario del aeropuerto y provisión de espacio de oficina para sus representantes;
- Control de carga, mensajería y telecomunicaciones;
- Manipulación, almacenamiento y administración de dispositivos de carga;
- Cualquier otro servicio de supervisión antes, durante o después del vuelo, y cualquier otro servicio administrativo solicitado por el usuario del aeropuerto.

‘Handling’ de pasajeros

El manejo de pasajeros comprende cualquier tipo de asistencia a los pasajeros que llegan, salen o transitan, incluyendo la comprobación de billetes y documentos de viaje, el registro de equipaje y su transporte a la zona embarque.

Manejo de equipaje

La gestión del equipaje comprende la manipulación del equipaje en el área de facturación, su facturación, su preparación para la salida, su carga y descarga desde los dispositivos diseñados para moverlo desde la aeronave hasta la zona de clasificación y viceversa, así como el transporte del equipaje desde el Área de facturación al área de recuperación de equipaje.

Manejo de carga y correo

- Para el transporte de mercancías: manejo físico del flete de exportación, transferencia e importación, manejo de documentos relacionados, procedimientos aduaneros y aplicación de cualquier procedimiento de seguridad acordado entre las partes o requerido por las circunstancias;
- Para el correo: Manejo físico del correo entrante y saliente, manejo de documentos relacionados y la implementación de cualquier procedimiento de seguridad acordado entre las partes o requerido por las circunstancias.

‘Handling’ de rampa

- Supervisar estacionamiento del avión en el parking y adecuar el entorno para su repostaje, tanto a la llegada como a la salida;
- Asistencia a la carga de aeronaves y provisión de dispositivos adecuados;
- Comunicación entre la aeronave y el proveedor de servicios;
- La carga y descarga de la aeronave, incluida la provisión y operación de medios adecuados, así como el transporte de tripulantes y pasajeros entre la aeronave y la terminal, y el transporte de equipaje entre la aeronave y la terminal;
- La provisión y operación de unidades apropiadas para el arranque del motor;
- El traslado de la aeronave a la llegada y salida, así como la provisión y uso de dispositivos adecuados;
- El transporte, carga y descarga de la aeronave de ‘catering’.

Manejo de combustible y aceite

- La organización y ejecución de operaciones de abastecimiento de combustible, incluyendo el almacenamiento de combustible y el control de la calidad y cantidad de entregas de combustible;
- La reposición de aceite y otros fluidos

Mantenimiento de aeronave

- Servicios de mantenimiento de rutina realizados antes del vuelo;
- Servicios no rutinarios solicitados por el usuario del aeropuerto;
- La provisión y administración de repuestos y equipo adecuado;
- La solicitud o reserva de un espacio de estacionamiento y / o hangar adecuado

Operaciones de vuelo y administración de la tripulación

- Preparación del vuelo en el aeropuerto de salida o en cualquier otro punto;
- Asistencia en vuelo, incluida la reexpedición si es necesario;
- Actividades posteriores al vuelo;
- Administración de la tripulación.

Servicios de 'Catering'

- Enlace con proveedores y gestión administrativa;
- Almacenamiento de alimentos y bebidas y del equipo necesario para su preparación;
- Limpieza de este equipo;
- Preparación y entrega de equipos, así como de bar y alimentos.

Estas actividades son a veces proporcionadas por un solo agente de 'Handling', o cada una de ellas puede ser responsabilidad de diferentes autoridades. Esta situación, es la raíz de la naturaleza compleja de las operaciones de 'Handling'. Por ejemplo, algunos servicios como catering, mantenimiento de aeronaves o abastecimiento de combustible pueden ser proporcionados por empresas externas, mientras que el embarque de pasajeros puede ser gestionado por la propia aerolínea o por agentes de 'Handling' independientes, y el proceso de rampa puede ser manejado por la propia compañía del Aeropuerto. Una aerolínea también puede utilizar diferentes combinaciones de proveedores internos y externos en diferentes aeropuertos.



Figura 2 - Proceso secuencial del 'Turnaround'

Las operaciones de 'Handling' en aeropuertos pueden ser realizadas por tres proveedores de servicios, diferentes. Estos son:

- Terceros u Operadores independientes: Operadores independientes, se refiere a cualquier proveedor de 'Handling' de tierra diferente al de la compañía que lo contrata. Los procesos de terceros són realizados por empresas de 'Handling' especializadas en operaciones de manipulación en tierra. Los operadores de terceros pueden operar en cualquier aeropuerto durante sus concesiones siempre que cumplan con los estándares acordados.

Los cambios fundamentales en el negocio aeroportuario desde el concepto tradicional de monopolio (proporcionando todos los servicios el mismo agente) para enfocarse en el negocio principal, provocaron la aparición de nuevos manejadores globales de terceros como Menzies, Globeground o Aviapartner. Estas empresas tienen su propio equipo de manipulación aparte de las aerolíneas, así como los aeropuertos. Este equipo constituye la mayor parte del capital de los activos de las empresas de 'Handling'.

- Autoasistencia: La autoasistencia se refiere a la situación en la que un usuario del aeropuerto proporciona directamente una o más categorías de servicios de asistencia en tierra y no realiza ningún contrato de terceros para la prestación de tales servicios. En la autoasistencia, la conducción de aeronaves es realizada por la propia línea aérea en lugar de otras partes.
- 'Airport Handler': Un aeropuerto puede servir a la aerolínea como agente de 'Handling', además de ser la interfaz donde aterrizan y despegan los aviones. Los aeropuertos que participan en los servicios de handling pueden aprovechar el hecho de estar a cargo de todas las actividades (equipamiento, puertas, stands, franjas horarias, etc.) para realizar el manejo de aeronaves y pasajeros. Especialmente en Europa, los operadores aeroportuarios que llevan a cabo la asistencia en tierra se han vuelto muy profesionales desde 1990.

Anteriormente, la estructura monopolística de la asistencia en tierra en los aeropuertos, especialmente en aquellos donde el operador del aeropuerto realizaba el 'Handling', resultó en un desafío para que los operadores individuales entraran al mercado. Tras la liberalización del mercado de la manipulación por la aplicación del Consejo Europeo, la estructura del mercado cambió ligeramente y se regula la entrada de terceros.

En la figura siguiente se puede observar que las cajas están coloreadas de distintos tonos. No se trata de una elección aleatoria sino que tiene un significado. Las actividades se han dividido en grupos: flujo de equipajes y mercancías (marrón), flujo de pasajeros y tripulación (naranja), flujo de cabina (amarillo), flujo de combustible (verde), y operaciones técnicas de la aeronave (morado). Esos grupos representan las actividades que deben realizarse durante el periodo de servicio en tierra. También se pueden encontrar cajas azules pero no forman parte del 'ground handling' tal, como se puede concluir con la definición del mismo. Tan solo se han incluido para ilustrar más claramente lo que sucede desde la llegada del avión hasta su salida.



Figura 3 - Diagrama de actividades del 'turnaround'

El tiempo en el que se realizan las actividades del 'ground handling' puede estar referido al momento de llegada o salida del avión indistintamente. En la figura 3.1 se puede ver esto con facilidad ya que dependiendo de qué referencia se tome las actividades se encontrarán a la izquierda o la derecha de la línea morada vertical. Las actividades a la izquierda son las que se planifican según la hora de llegada, mientras que las de la derecha toman la salida como referencia. Entre los dos tipos de actividades descritas arriba (según la referencia tomada) puede hacer un tiempo de inactividad, que en caso de existir puede ser más o menos largo según el tiempo que permanezca el avión en el aeropuerto. Si la estancia es lo suficientemente corta todas las actividades pueden realizarse consecutivamente, sin ningún momento de inactividad.

La duración del 'turnaround' depende de los siguientes factores:

1. *Tamaño del avión: cuanto mayor sea el avión más durará el 'turnaround'. Existe un tiempo mínimo (en torno a veinte minutos) que es el tiempo que tardan en enfriarse los frenos aproximadamente.*
2. *Tipo de vuelo: los vuelos de corta distancia suelen operar con horarios muy ajustados, mientras que los de larga distancia suele contar con unos márgenes de tiempo más amplios.*
3. *Número de pasajeros y cantidad de mercancía.*
4. *Estrategia de la aerolínea: algunas aerolíneas deciden dejar un tiempo de inactividad entre tareas para asegurarse de que pueden cumplir con el horario previsto.*

A la llegada del avión, las primeras actividades que se deben realizar son el desembarque de tripulación y pasajeros, y la descarga de equipajes, siempre y cuando se puedan llevar a cabo en condiciones seguras. Es fundamental garantizar que los pasajeros sufran el mínimo retraso posible de manera que se minimice el riesgo de perder hipotéticos vuelos de conexión.

Una vez que se termine el desembarque y descarga, si otro avión necesita hacer uso del puesto de estacionamiento, y siempre y cuando la siguiente salida no tenga un margen horario muy ajustado, se puede llevar al primer avión a un estacionamiento remoto.

La descarga/carga de mercancías se puede realizar más o menos rápido en función de cómo de urgente sea y de la disponibilidad de recursos para hacerlo en la ubicación deseada. Del mantenimiento de la aeronave se encarga la propia aerolínea y se puede realizar en el lugar donde se encuentre estacionada. La limpieza y saneamiento también pueden realizarse in situ, con la hora de llegada como referencia normalmente, para tener el avión limpio cuanto antes, pero en realidad también pueden referirse a la salida. Es tan sólo una cuestión de encontrar el momento oportuno para conseguir que sea más económico y que no interfiera con otras actividades.

A medida que se aproxime la hora de salida hay que realizar otras actividades. Si el avión está estacionado en un puesto remoto puede ser acercado a un puesto en contacto si es necesario. En general el reabastecimiento se realiza cuando la aerolínea lo requiere. Un vez que se ha completado el reabastecimiento, carga de equipajes y mercancías, y embarque, el avión está totalmente listo para abandonar la plataforma. El 'push-back' se realiza una vez que el piloto recibe el permiso (clearance) de la torre de control para despegar.

El 'ground handling' se caracteriza por la diversidad de actividades que abarca, al igual que el amplio rango de equipos y vehículos que se requiere. La complejidad es otra de sus características destacadas, al requerir que se realicen distintas actividades en paralelo y/o secuencialmente.

1.2.- Retrasos de 'Handling'

Cada factor que tiene un impacto en el rendimiento de la aerolínea requiere un aviso especial. El número de entregas de equipaje no entregados o tardíos, el check-in y el embarque con retraso y el bajo rendimiento en tiempo debido a las operaciones en tierra son algunos de los indicadores de desempeño de las aerolíneas que están estrechamente relacionados con las operaciones de 'Handling'. El nivel de servicio de las operaciones de asistencia en tierra afecta a los pasajeros, a otras compañías aéreas, así como a todo el funcionamiento del aeropuerto en relación con el lado terrestre (por ejemplo, mostradores de facturación, entrega de equipaje) y al aire (por ejemplo, Rampa, tiempo de respuesta máximo ...). Por lo tanto, se puede concluir que las operaciones de 'Handling' no sólo pueden dar lugar a retrasos de los vuelos, sino que también tienen un impacto en las otras actuaciones de una aerolínea relacionados con las experiencias de los pasajeros. [4]

Cada año, las aerolíneas han experimentado retrasos adicionales. El porcentaje de cada razón de retraso puede diferir de un año a otro. Hasta ahora, una de las principales preocupaciones de las aerolíneas eran las demoras ATFM (Air Traffic Flow Management). El espacio aéreo limitado en el aeropuerto y en ruta y el aumento de las operaciones aéreas en ciertas rutas habían afectado a todas las redes de transporte. Los intentos de resolver estos problemas por unidades relacionadas, como EUROCONTROL y ACI, resultaron en un resultado positivo, y los retrasos debidos a ATFM y ATC han comenzado a disminuir. Las operaciones aeroportuarias y de manipulación que han sido desatendidas en el pasado han comenzado a atraer la atención de las aerolíneas.

Dado el hecho de que un coste de retardo de un minuto en tierra resulta en aproximadamente 50 €, mientras que en el aire es de 70 € (Según datos de Eurocontrol) es de suma importancia que las aerolíneas encuentren formas de hacer frente a los retrasos. El coste del combustible, el coste de la tripulación de vuelo, el coste de la aeronave arrendada, los gastos aeroportuarios y también los costes no medidos (por ejemplo, las quejas de los clientes y el coste de deslealtad) son algunos de los ejemplos que las compañías aéreas tienen que afrontar como consecuencia de un aumento de los retrasos.

Los retrasos en la cadena de 'Handling' no sólo provocan impactos en la calidad del servicio experimentado por los pasajeros, sino que también afectan la eficiencia operativa y, en consecuencia, los costes de la aerolínea. Los retrasos resultantes de la asistencia en tierra constituyen uno de los costes más elevados de las compañías aéreas, a pesar de que los retrasos relacionados con la gestión son una forma más barata y más sencilla de reducir los retrasos de salida y, en consecuencia la dificultad de reducir otros motivos de retrasos, como las condiciones meteorológicas y el control del tráfico aéreo (ATC).

Los retrasos en las aerolíneas se deben a varias razones, y el 'Handling' en tierra es sólo uno de ellos.

Existen muchas interrupciones potenciales en la cadena de acciones y estos pueden ser inhibidos por las llegadas tardías y los retrasos en el control aeroportuario y aéreo (ATC) que acortan el tiempo disponible para los cambios. El problema característico de manejar los retrasos es que hay muchos hechos que influyen en el proceso de 'Handling', así como la falta de procedimientos para monitorear la interacción entre estos hechos en un período de 'turnaround' limitado.

La eficiencia global del transporte aéreo depende, por lo tanto, del buen funcionamiento de todos los elementos de la cadena de suministro de la aviación. Sin embargo, las aerolíneas forman parte de un sector de servicios de transporte más amplio que abarca, entre otros, aeropuertos, servicios de navegación aérea, aeromotores y fabricantes de motores aéreos, así como sofisticados sistemas de información y billetes.

La cadena de valor de Michael Porter es un buen ejemplo de mostrar la localización de la compañía de 'Handling' en las operaciones de la aerolínea. La cadena de valor de Porter se basa en la visión del proceso de las organizaciones como un sistema, compuesto por subsistemas con entradas, procesos de transformación y resultados. Cada actividad en diferentes niveles agrega cierto valor al producto final.

Porter's value chain

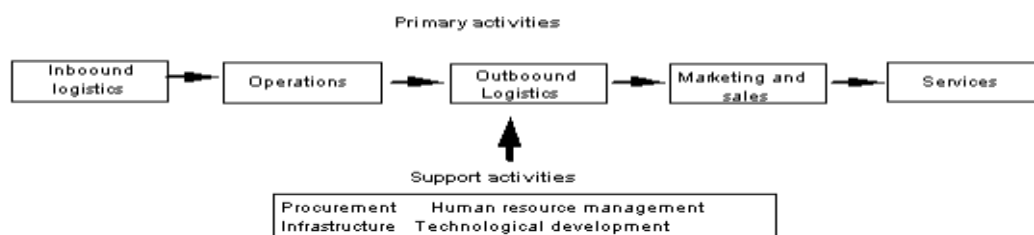


Figura 4 - Cadena de valor de Porter

Según Porter (1985),

Las principales actividades en la cadena de valor son:

- 1. Logística de entrada:** implica relaciones con proveedores e incluye todas las actividades necesarias para recibir, almacenar y difundir los insumos.
- 2. Operaciones** - son todas las actividades necesarias para transformar insumos en productos (productos y servicios).
- 3. Logística de salida:** incluye todas las actividades necesarias para recopilar, almacenar y distribuir la salida.
- 4. Marketing y Ventas** - las actividades informan a los compradores sobre los productos y servicios, inducen a los compradores a comprarlos y facilitan su compra.
- 5. Servicio** - incluye todas las actividades necesarias para mantener el producto o servicio funcionando eficazmente para el comprador después de que se venda y sea entregado.

Las actividades secundarias son:

- 1. Adquisición** - es la adquisición de insumos, o recursos, para la empresa.
- 2. Gestión de recursos humanos:** consiste en todas las actividades relacionadas con la contratación, capacitación, desarrollo, compensación y (si es necesario) despido o despido de personal.
- 3. Desarrollo Tecnológico** - se refiere al equipo, hardware, software, procedimientos y conocimientos técnicos que se llevan a cabo en la transformación de los insumos en productos por parte de la empresa.
- 4. Infraestructura** - sirve a las necesidades de la empresa y vincula sus diversas partes, que consiste en funciones o departamentos tales como contabilidad, legal, finanzas, planificación, relaciones públicas, relaciones gubernamentales, garantía de calidad y gestión general.

A menor escala, si la cadena de valor de Porter se aplica al negocio de la aerolínea, los insumos pueden asignarse como equipaje y pasajeros, mientras que el 'Handling' en tierra puede asignarse como procesos de "operación" y "logística de salida" para ser el 'output' un vuelo de salida a tiempo.

El nivel de eficiencia de los estándares que se llevan a cabo en las actividades de 'Handling' en cada nivel de la cadena determina la calidad global del servicio suministrado al pasajero al final.

1.3.- Definición del 'turnaround' de las Aeronaves

El cambio de aeronave, que es el núcleo del negocio de manipulación, puede definirse como el proceso de preparar el avión para el siguiente vuelo. Este proceso se logra mediante el trabajo simultáneo de diferentes departamentos operativos. [7]

El proceso de 'turnaround' se establece en un período de tiempo definido, bajo cualquier circunstancia dada, y debe hacerse sin perder tiempo ni recursos. El tiempo de respuesta se refiere al tiempo transcurrido entre el 'on-block' y el 'off-block' de la aeronave. Durante este período, muchas tareas se llevan a cabo simultáneamente y en relación entre sí.

El tiempo de respuesta de las aeronaves tiene que estar en los límites de tiempo de tierra estándar, que es alrededor de 45-90 minutos dependiendo del tipo de la aeronave. Las diferencias en los tiempos en tierra y, en consecuencia, las diferencias en los tiempos de respuesta de las aeronaves entre los principales operadores y los bajos costes dependen de las estrategias adoptadas por las aerolíneas.

La eficiencia del proceso de 'Ground Handling' depende principalmente de cuestiones como la disponibilidad de personal, la programación, la cantidad de equipo, el posicionamiento, la adaptabilidad a los picos de tráfico, las tareas de trabajo, la meteorología, las condiciones, la capacidad, el rendimiento y la compatibilidad del personal y del equipo de asistencia en tierra. Por otra parte, su eficiencia depende de cómo todas estas tareas y condiciones simultáneas se gestionan de manera eficaz y coordinada.

En este aspecto, se puede inferir que el proceso de reacondicionamiento de las aeronaves es muy complejo. Cuando el avión llega a su puesto de estacionamiento, muchos departamentos operativos diferentes trabajan juntos para preparar el avión. Tales como:

- Los servicios de pasajeros son responsables de que los pasajeros sean inspeccionados y embarcados en la salida.
- Los servicios de rampas son responsables de la carga de las equipajes y acondicionamiento de las aeronaves.
- La empresa de limpieza se encarga de limpiar la cabina.
- La compañía de repostaje de fuel es responsable de entregar el combustible solicitado a tiempo.
- La compañía de Catering es responsable de subir las galeras nuevas y quitar las vacías.
- La Autoridad Aeroportuaria es responsable de proveer servicio de puente, puertas de embarque y servicios personalizados (a menos que sea un proveedor de 'Handling').
- El Departamento de Carga es responsable de preparar los documentos relacionados con la carga y entregar la carga al departamento de rampa.
- El Departamento de Bomberos es responsable de estar en la posición de estacionamiento si hay embarque, y el abastecimiento de combustible se realiza al mismo tiempo.
- Operaciones de Vuelo es responsable de preparar la carga y el balance, y entregarlo a tiempo.

En este sentido, todos estos departamentos trabajan juntos de manera interactiva durante los cambios de aeronave para llevar a cabo los diferentes procesos que se muestran en la siguiente figura.

Como se muestra en esta figura, la mayoría de estas tareas dependen o se basan en otra tarea. Por ejemplo, antes de que finalice la limpieza, los pasajeros no pueden ser abordados, o antes del embarque de la tripulación, el aprovisionamiento de combustible no se puede completar.

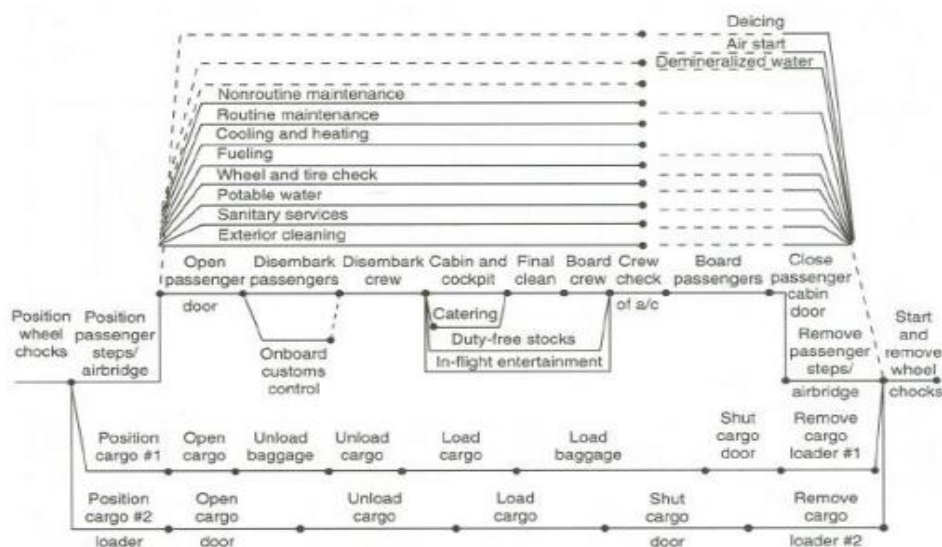


Figura 5 - Representación gráfica del 'Turnaround'

La mayoría de las tareas de 'turnaround' son secuenciadas. La limpieza tiene que esperar para el desembarque, el embarque tiene que esperar a que la tripulación llegue y la limpieza sea completada, el 'Duty free' tiene que esperar a que la tripulación de cabina esté a bordo, el abastecimiento de combustible tiene que ser iniciado después del desembarque y debe ser terminado antes del embarque de Pasajeros, etc.

En consecuencia, si la tripulación llega tarde, la carga de 'Duty Free' será tarde; Si el 'Duty free' se atrasa, entonces el embarque será tarde; Si la limpieza tarda más de lo programado, el embarque será atrasa; Si el embarque se atrasa, el proceso de sentar a los pasajeros será tarde; Si la carga de equipaje se retrasa, el cierre de las puertas de los aviones será tarde, y como resultado, el 'Push back' y la salida de la aeronave será retrasará.

La sincronización de las actividades de rampa proporciona eficacia en las operaciones y también a otras tareas en la cadena que se realizará sin ningún defecto. Para proporcionar esta sincronización, hay una persona responsable en rampa a la que se puede llamar 'redcap', personal de operaciones (coordinador de rampa) o representante de la aerolínea. Esta persona actúa como un jefe de orquesta y controla todas las tareas al mismo tiempo. Las funciones del personal de operación o coordinador de rampa pueden variar de acuerdo con la estrategia de la aerolínea y el aeropuerto dedicado al tema. El 'turnaround' se realiza en casi una hora, y las consecuencias de las operaciones problemáticas tienen un gran impacto en el resto de las operaciones que el aeropuerto y la línea aérea tienen que seguir.

La eficiencia de una operación de 'turnaround' se define como la capacidad de una aerolínea para ejecutar los servicios de acondicionamiento de aeronaves requeridos dentro del tiempo de servicio disponible y para asegurarse una salida de vuelo puntual. Pero no es fácil realizar un 'turnaround' eficiente debido a muchas circunstancias que pueden ocurrir durante este período de tiempo.

Otro punto que vale la pena mencionar con respecto a las operaciones de 'turnaround' es que hay muchos tipos de equipos móviles para la asistencia en tierra y muchas tareas periféricas tanto a nivel de pista como a nivel superior en un espacio y tiempo limitados, como se ve en la siguiente figura. Estas operaciones requieren la concentración exacta y densa de las tripulaciones cuando están maniobrando y colocando el equipo para realizar el proceso de 'Handling' del avión. [6]

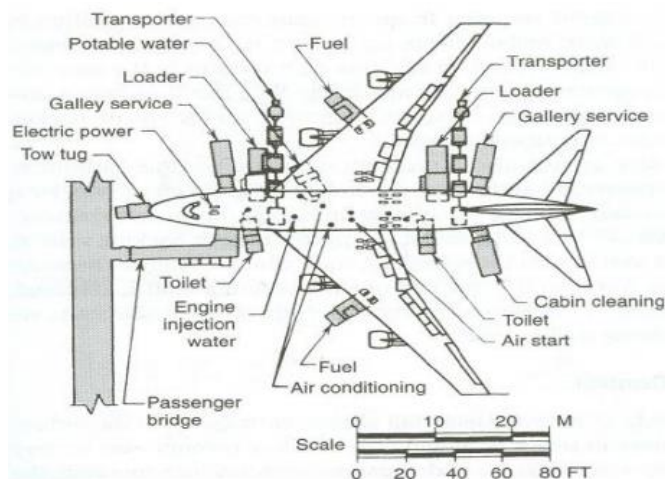


Figura 6 - Representación de la colocación de los equipos de 'Handling'

Caso 3: La aeronave llega a tiempo y sale antes de lo planeado, ni siquiera incluyendo el tiempo de amortiguación a menos que esté en el límite del 'SLOT'. Este caso depende de la cantidad de carga y número de pasajeros en la aeronave, y también el rendimiento del equipo de tierra y el trabajo en equipo eficiente de los diferentes departamentos. Aquí, la situación es $ATD < STD$

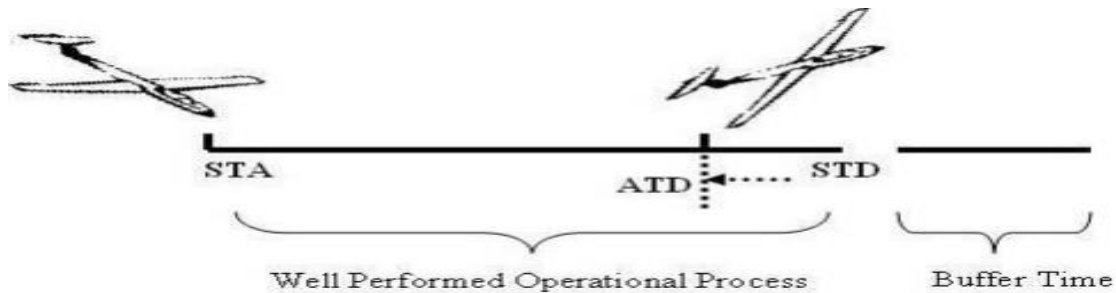


Figura 9 - Representación llegada en hora y salida adelantada

Caso 4: Las aeronaves llegan tarde y se utiliza tiempo de amortiguación para absorber el retraso de llegada por el proceso planeado de 'turnaround' en tierra. Aunque se espera que sea una operación estándar normal, algunas aerolíneas, como las 'Low Cost', presionan a los Operadores de 'Handling' para que completen el 'turnaround' antes del tiempo de amortiguación. La ecuación es $ATD = STD + Buffer Time$

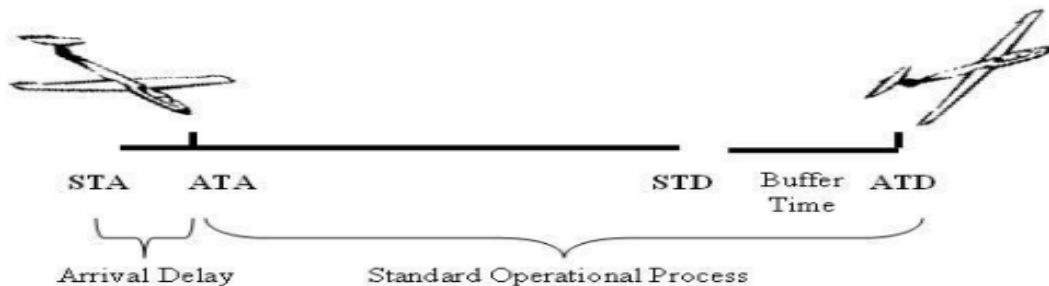


Figura 10 - Representación llegada tardía y salida en hora (Buffer time)

Caso 5: Las aeronaves llegan tarde y salen a tiempo, sin incluir el tiempo de amortiguación, debido al eficiente manejo en tierra. En esta situación, el manejo de tierra eficiente cubre el retraso de llegada y también los retrasos reaccionarios. La ecuación es $ATD = STD$



Figura 11 - Representación llegada tardía y salida en hora (Durante Buffer time)

Caso 6: Las aeronaves llegan tarde y salen tarde al superar el tiempo de amortiguación debido a la manipulación en tierra o cualquier otra razón. Adicional al retraso de llegada, si el manejo en tierra es ineficiente, esto aumenta el retraso de salida que puede causar enormes costes tanto a la aerolínea como a la empresa de 'Handling'. $ATD = STD + \text{Buffer Time} + \text{Delay}$

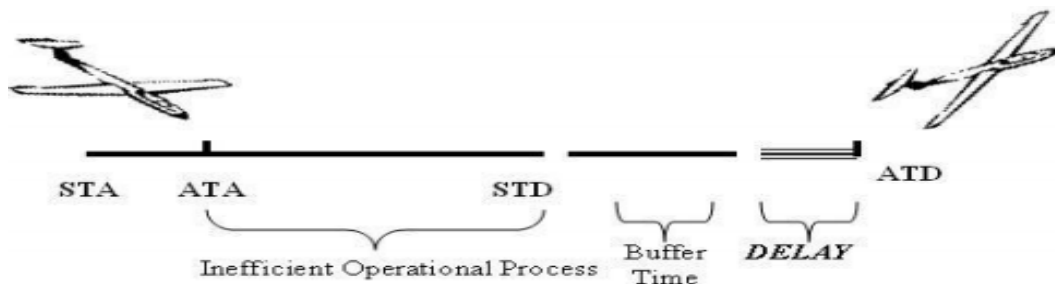


Figura 12 - Representación llegada tardía y salida tardía

Caso 7: La aeronave llega a tiempo y sale tarde debido a una manipulación en tierra ineficaz o cualquier otra razón. Esta situación es la peor. Los operadores de tierra, con la falta de equipo y personal, en las temporadas de pico pueden ser una de las razones de este tipo de retraso.

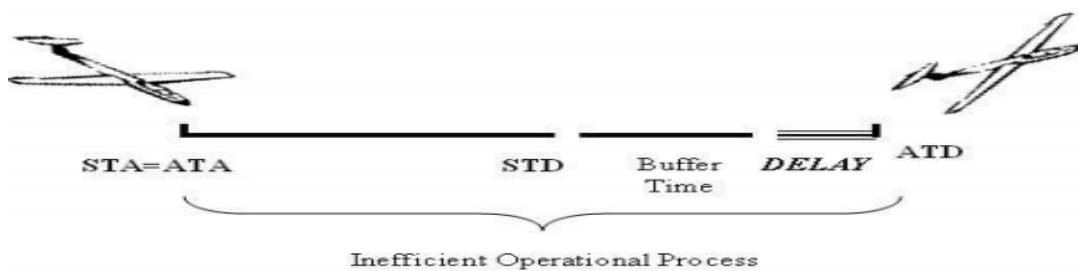


Figura 13 - Representación llegada en hora y salida tardía

Mediante el uso del 'Buffer time', el agente de 'Handling' tiene capacidad adicional para gestionar los posibles retrasos. Por otro lado, existe un equilibrio óptimo entre el tiempo de amortiguación programado y el coste esperado del sistema. Cuanto más tiempo un avión permanece en tierra, según el tipo y la clase, más aumentarán los gastos. En cada minuto adicional en tierra, la aerolínea tiene que pagar una cantidad adicional a las autoridades aeroportuarias: el precio unitario se multiplica por el peso del avión por cada minuto si permanece en el puesto de estacionamiento excediendo el tiempo planeado.

Las aerolíneas programan sus vuelos cerca del horario mínimo, aunque son conscientes de que tendrán retrasos en ciertas rutas. Las aerolíneas pueden planificar un horario más largo para que su red entera no se anuncie como retrasada, y también planificar un tiempo de programación más largo puede dar suficiente tiempo al Operador para el 'turnaround'. Aunque las aerolíneas pueden utilizar el tiempo de amortiguación en los horarios de vuelo para controlar la puntualidad del horario, disminuyen la utilización de aeronaves y aumentan los costes de personal. La reducción de cinco minutos de este tiempo de amortiguación supondría alrededor de 1.000 millones de euros ahorrados, en una mejor utilización de los recursos de las líneas aéreas y los aeropuertos. Por esta razón, en la mayoría de los aeropuertos ocupados, los Operadores tienen que lograr el 'turnaround' de la aeronave en tiempos mínimos para servir a las aerolíneas.

Cada tipo de aeronave requiere un cierto tiempo de ‘turnaround’ según el tipo de destino, como el de largo recorrido o de corto recorrido, y el tipo de aerolínea, como la de bajo coste o la ‘full carrier’. En la siguiente figura, se muestra el tiempo de respuesta aproximado de un B-747 y las tareas principales. La duración de cada tarea se muestra en la figura.

La duración del cambio depende no sólo del tipo de aeronave y del nivel de servicio de la aerolínea, sino también del número de pasajeros, de la cantidad de carga, de la eficiencia de la tripulación de tierra y de las comunicaciones exactas entre los departamentos. Hay muchas tareas que se pueden manipular en un período de tiempo más corto que el tiempo estándar si hay condiciones convenientes durante el cambio.

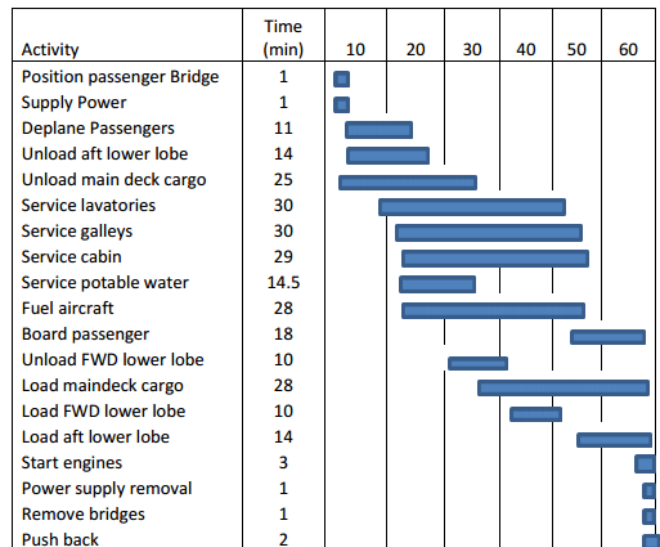


Figura 14 - Planificación (en diagrama de barras) de tiempos de ‘Turnaround’

1.5.- El proceso ‘Turnaround’

El proceso de ‘turnaround’ de las aeronaves comienza realmente un tiempo antes del aterrizaje de la aeronave. La preparación y el proceso de acondicionamiento pueden ser examinados en 3 etapas: antes de la llegada, ‘On the ground’ y después de la salida. Pese a eso, nosotros nos centraremos en las dos primeras etapas. La siguiente parte mencionará pequeños detalles y fallos que comúnmente experimentan los agentes de manipulación en todas las etapas del proceso de ‘turnaround’.

1.5.1.- Antes de la llegada

Las aeronaves requieren pre-provisión antes de su llegada según el tipo de vuelo, así como la aeronave, la cantidad de pasajeros y la carga. Los aeropuertos, especialmente los congestionados, manejan muchos aviones durante las horas punta. Por lo general, las llegadas consecutivas hacen que sea más difícil para los Operadores satisfacer la demanda en las horas punta. La preparación previa del agente de ‘Handling’ ayuda a disminuir el caos durante las horas punta. Las preparaciones que el agente de manipulación tiene que considerar antes de la llegada de las aeronaves son:

Comprobación de datos de vuelo

Hay algunos datos importantes sobre los vuelos que deben ser comprobados antes de la llegada, así como la salida de la aeronave. Los permisos de vuelo, las franjas horarias y el horario del plan de vuelo de la aerolínea son los más importantes a considerar antes de la salida del vuelo.

Los permisos otorgados por las autoridades relacionadas en vuelo fecha y hora, también tienen que ser controlados. Los demás datos que también deben comprobarse es el de la franja horaria del aeropuerto (SLOT), especialmente en caso de desviación a la hora de llegada o salida. Por ejemplo, si hay un retraso a la llegada al aeropuerto donde se tiene una solicitud de franja horaria, la información del retraso debe ser declarada a la autoridad aeroportuaria, y se solicita una nueva franja horaria de acuerdo con la nueva hora de llegada del avión. Además, un nuevo tiempo de 'SLOT' del Control de Tráfico Aéreo (ATC) tiene que ser obtenido por los expedidores de la línea aérea o por el personal de la operación del agente de 'Handling'.

Como se muestra en el siguiente ejemplo, la revisión 'SLOT' se solicita desde Bruselas con un mensaje de 'SLOT'. Este mensaje solicita la nueva hora de salida. Si el personal de operación olvida renovar el tiempo de la ranura con este mensaje, la aeronave puede permanecer en el suelo durante horas. El mensaje se envía a EUROCONTROL para solicitar un nuevo tiempo de despegue calculado (NEWCTOT). En este mensaje se declara el motivo del retraso y se solicita la disponibilidad de nuevo tiempo de despegue de EUROCONTROL en el formato del ejemplo:

BRUEA7X 031221	→	Dirección EUROCONTROL
TITLE SRM	→	Mensaje Slot de Revisión
BRUEA7X 031221	→	Dirección EUROCONTROL
ARCIDUA826	→	Identificación de llamadas de aeronaves
IFPLID BB67317757	→	Identificación del plan de vuelo
ADEP LTAC	→	Aeropuerto de salida
ADES LOWW	→	Aeropuerto de destino
EOBT 060203	→	Fecha estimada de operación
NEWCTOT 1251	→	Nuevo tiempo de despegue calculado
REGUL LOWWA03M	→	El motivo de la revisión de SLOT
TAXITIME 0006	→	Duración del rodaje
REGCAUSE GA 87	→	El código de la Causa de demora
EOBT 1245	→	Tiempo de 'Block' estimado

Figura 15 - Ejemplo. Revisión de SLOT

Comprobación de los datos del avión

Todas las aerolíneas tienen información sobre las aeronaves en su flota en un documento denominado Manual de Operación Terrestre (GOM). Antes de la llegada de la aeronave, las instrucciones de carga, la configuración de la aeronave y las limitaciones de la aeronave se verifican desde el GOM por el agente de 'Handling'.

El GOM incluye información como las dimensiones de las bodegas (equipaje y carga), las configuraciones del asiento del pasajero, los pesos operativos y los límites del punto del centro de gravedad. Esta información ayuda a lograr la mejor carga de la aeronave. El punto de centro de gravedad de un avión puede diferir según su tipo.

Por ejemplo, el punto de gravedad en los aviones de tipo MD está hacia popa, lo que significa que la mayoría de las bolsas tienen que cargarse en la parte trasera de la aeronave para obtener el mejor equilibrio, mientras que Airbus 360 requiere carga frontal.

Si el personal de operación no comprueba esta información y si la carga se realiza de manera incorrecta, el avión estará fuera de recorte y será necesario descargar el avión y volver a cargarlo, lo que se traduce en un promedio de 30 minutos de retraso, dependiendo de la cantidad de la carga. Hoy en día, la mayoría de las aerolíneas utilizan sistemas computarizados para hojas de carga, lo que facilita al personal de operación preparar los datos de carga para el vuelo relacionado. En estos sistemas, es más fácil ver los fallos de carga y las limitaciones de carga de la aeronave en el ordenador. Si se trata de una hoja de carga computarizada, es de suma importancia conocer estos detalles con el fin de poder preparar la carga y rellenar la hoja lo más rápidamente posible.

Evaluación de los mensajes de vuelo

La evaluación de los mensajes significa entender la necesidad del mensaje y distribuirlo a las unidades relacionadas. Esta acción es necesaria para que todas las partes estén listas para las operaciones de la aeronave. Cada mensaje incluye información diferente relacionada con diferentes agentes. Por ejemplo, un mensaje de llegada estimado (EA) es importante para la autoridad aeroportuaria para arreglar estacionamientos y puertas, para actualizar las tarjetas de información en el edificio de la terminal, etc., mientras que un mensaje de carga es importante para el departamento de rampa para preparar el equipo y personal.

Frecuencia de radio

Las aeronaves que están cerca de la aproximación final o en la misma aproximación final, son capaces de comunicarse con agentes de 'Handling' a través de frecuencias de radio (VHF).

A través de VHF, el piloto puede informar al personal de tierra acerca de lo que necesitan en la posición de estacionamiento e intercambiar información operacional. Esta frecuencia se puede utilizar para reducir el tiempo de respuesta. Por ejemplo, antes del aterrizaje de la aeronave, los datos de la hoja de carga pueden solicitarse a la cabina, que se recibe tradicionalmente después del aterrizaje. De esta manera, la hoja de carga se puede preparar antes y pasar al capitán para la salida en un período de tiempo reducido. Actualmente, 'Ryanair', una compañía de bajo coste, utiliza esta herramienta para que el Operador comience el embarque de sus pasajeros justo antes de su aterrizaje.

Llegada a la posición de estacionamiento

Un cierto número de personal de diferentes departamentos tiene que estar en la posición de aparcamiento antes del aterrizaje del avión. El personal de operación comprueba si todo el equipo y la tripulación están listos o no, y se somete a un 'briefing' (informe) con el jefe de rampa sobre el proceso de carga y descarga; El supervisor de la rampa controla el número de trabajadores y el equipo de la rampa, y se somete a una información sobre la carga / descarga; El personal de tránsito prepara la documentación para pasársela a la tripulación de cabina.

Posicionamiento del equipo

Cada tipo de aeronave requiere diferentes equipos dependiendo de su diseño. Según el tipo de avión, una cantidad adecuada y suficiente de equipo se lleva a la posición de estacionamiento. Por ejemplo, si el tipo de avión requiere dispositivos de carga unitaria (ULD) para cargar, se debe traer equipo relacionado como carros, cargadores o montacargas. Es crucial traer el equipo comprobado (si está funcionando correctamente o no) a la posición del parking. Si se trata de carga a granel de acuerdo con los requisitos de carga a granel deben llevarse los tractores, un número suficiente de carros y transportadores.

Instrucciones

Cuando un representante de la aerolínea no está a cargo del proceso de 'turnaround' de la aeronave, el supervisor de rampa está en comunicación con el representante de la aerolínea para administrar el proceso. La información incluye el intercambio de información final con los representantes de las aerolíneas y unidades relacionadas. A veces, los representantes pueden solicitar servicios adicionales como un autobús extra, servicio VIP o algunas diferencias en el servicio planificado. Por ejemplo, en lugar de una limpieza completa, pueden solicitar una limpieza en tránsito o una brigada de bomberos, etc. Después del 'briefing', si hay este tipo de requisitos, se pasan a las unidades relacionadas antes de la llegada.

1.5.2.- 'On the ground'

Desembarque de pasajeros

Los pasajeros son desembarcados de tres maneras: a través de las pasarelas (fingers), a través de los autobuses en las posiciones remotas, o caminando de avión a terminal (sólo en estacionamiento cerca de terminal).

Si la aeronave está estacionada en una posición remota, los autobuses tienen que estar en el puesto de estacionamiento para cumplir con el vuelo a su llegada. En los puestos de estacionamiento remotos, el desembarque rápido requiere suficientes autobuses para los pasajeros. El número de autobuses en el desembarque o embarque de pasajeros se decide en los acuerdos de nivel de servicio con la línea aérea. Las compañías aéreas tienen diferentes opciones para el número de autobuses. Esta decisión depende de cómo la línea aérea quisiera servir a sus clientes. Por ejemplo, si bien un transportista de servicios completos puede preferir dos autobuses para 80 pasajeros, una compañía 'charter' o de bajo coste puede preferir un autobús para el mismo número de pasajeros. La responsabilidad de la asistencia en tierra es proporcionar los números solicitados, como se acordó con la compañía aérea en el acuerdo de nivel de servicio, y hacer que los autobuses estén listos en la posición.

Para las operaciones de puente, un operador de puente aéreo va al panel de control del puente para colocar el puente aéreo. En las pasarelas aéreas, que en su mayoría están bajo la dirección de las autoridades aeroportuarias, un operador de puente debe estar en el puesto de estacionamiento al menos en el momento en que llega el avión.

Limpieza

La limpieza de la aeronave es realizada por empresas de limpieza subcontratadas o por el agente de 'Handling'. Las aerolíneas notifican el número de personal de limpieza necesario en cada tipo de avión y el tipo de limpieza que prefieren en su contrato de servicio con la empresa de manipulación. La empresa de 'Handling' es responsable de cumplir los requisitos de los acuerdos de nivel de servicio, ya sea por sí mismo o a través del subcontratista de limpieza.

La limpieza de un avión puede variar dependiendo del tipo de vuelo. La limpieza de las aeronaves puede clasificarse como limpieza de tránsito, limpieza estándar y limpieza profunda. La limpieza estándar de los aviones a bordo incluye la limpieza de los inodoros, la restauración de tablas, la eliminación de la basura, limpieza de vacío, limpiar los pisos y asientos, limpiando las mesas de los asientos, limpiando y ordenando los bolsillos del asiento, arreglando los cinturones de seguridad, Y cubiertas de la cabeza, y limpiando las ventanas. En los vuelos de tránsito, la limpieza suele estar compuesta por la eliminación de la basura y la limpieza de los aseos, lo que lleva aproximadamente 15 minutos. Cuando la limpieza profunda se realiza con detergentes higiénicos y la limpieza tarda más.

La duración de la limpieza es importante porque hasta que no se haga la limpieza, no se puede iniciar el embarque. El número de personal de limpieza y el tipo de aeronave están estrechamente relacionados con la duración de la limpieza. Por ejemplo, en un modelo pequeño con un solo corredor de aviones como MDs o B737, tener equipos de limpieza adicionales no ayuda a acelerar la operación, sino que obstaculiza el flujo de trabajo en su interior. Por lo tanto, un número óptimo de personal de limpieza tiene que ser preestablecido, dependiendo del tipo de avión con respecto al acuerdo de nivel de servicio de la línea aérea.

Duty Free

Las compañías 'Duty Free' cargan productos para las ventas a bordo. La carga de 'duty free' se realiza con la ayuda de la tripulación de cabina, así como la correspondiente empresa. La carga comienza un tiempo después del desembarque de los pasajeros, y debe completarse antes de que los pasajeros lleguen al avión. En los vuelos que tienen carga de DF, el tiempo de respuesta es normalmente mayor porque operadores del avión y también azafatas son mantenidos ocupados por la empresa.

Abastecimiento / 'Catering'

El servicio de catering en el 'turnaround' incluye la remoción de las galeras vacías y su sustitución por las nuevas antes de que comience el embarque. Es crucial para el personal de operación organizar el flujo de trabajo a bordo para tener una carga más rápida. Específicamente, el movimiento dentro de los aviones de cuerpo pequeño/estrecho es muy difícil en un período de tiempo limitado. El uso de las puertas del avión entre catering, limpieza y 'duty free' tiene que ser muy eficiente.

Por lo general, las empresas externas, tales como la restauración y el abastecimiento de combustible se encuentran lejos de la zona de operaciones en tierra por lo que llegar a la zona de abastecimiento requiere un cierto tiempo. En los 'turnarounds' la falta de comida u ofertas adicionales de tripulación de cabina requiere más tiempo para completar la restauración, por lo que puede ser útil para comprobar el número de servicios de comidas frente al número de pasajeros para que no haya escasez de 'Catering'.

Esta concordancia cruzada debe ser realizada principalmente por la empresa de catering, y también como medida de precaución, la comprobación cruzada puede ser realizada por el personal de 'Handling' o por el representante de la aerolínea.

Carga / descarga de equipaje

La carga / descarga de la aeronave es la parte más importante del 'turnaround' de aeronave, que se realiza bajo la responsabilidad de dos unidades diferentes: operaciones y rampa.

Responsabilidad del departamento de operaciones: La responsabilidad de operaciones es planificar la carga / descarga considerando las reglas y los límites. Cada tipo de aeronave requiere ciertas formas de carga para un despegue seguro. La distribución de los pasajeros en la cubierta superior y la distribución de las bolsas y la carga en la cubierta inferior están planificadas por el personal de operaciones. El personal de operaciones considera los factores tales como la limitación de las bodegas, el centro de gravedad de la aeronave y la cantidad de carga útil (peso total de pasajeros, equipaje y carga) para el proceso de carga / descarga.

Responsabilidad del Departamento de Rampa: El proceso de carga / descarga requiere diferentes métodos según el tipo de aeronave. En la aeronave donde el centro de gravedad está en frente, la descarga tiene que comenzar desde el frente y seguir hacia la parte posterior. Para la aeronave donde el centro de gravedad está en la parte trasera, la carga tiene que partir desde el frente y seguir hacia atrás, para mantener la aeronave en equilibrio durante la carga / descarga. El personal de Rampa es responsable de considerar estas reglas durante el proceso de carga / descarga. La siguiente figura es un muy buen ejemplo de descarga incorrecta.



Tan pronto como el personal retiró la carga de la bodega delantera, la parte trasera del avión se derrumbó debido al hecho de que el avión estaba fuera de balance. La descarga de esta aeronave debería haber comenzado desde la bodega trasera para tener una descarga equilibrada y segura. ALITALIA 2012.

Figura 16 - Error en la descarga, al no respetar el balance de la aeronave.

Abastecimiento de combustible

El abastecimiento de combustible puede ser suministrado a la aeronave por los camiones de combustible o sistema de abastecimiento de hidrantes que se encuentra en cada puesto de estacionamiento. A pesar de que las compañías de combustible tienen los horarios de vuelo para servir a la aeronave, lo único que puede hacer la compañía de 'Handling' es llamar a la compañía de suministro de combustible a tiempo, e informarles sobre los cambios en el horario en caso de haber imprevistos. El personal de tierra se encarga de verificar si hay un camión de combustible en el puesto de estacionamiento cuando el avión llega al parking.

Anuncio de embarque

Una vez completada la preparación de la aeronave en tierra, se inicia el proceso de embarque. La iniciación del proceso de embarque requiere la comunicación entre la rampa, tráfico y la tripulación de cabina. Antes de iniciar el embarque, el personal de operación debe evaluar:

- Tiempo aproximado de finalización de los procesos de limpieza e información del tiempo de embarque a los servicios de pasajeros y de rampa: Información, los servicios de pasajeros anuncian el tiempo de embarque a los pasajeros, y los servicios de rampa organizan los autobuses para las posiciones de estacionamiento remoto;
- El número de autobuses que el agente de 'Handling' ha asignado para el vuelo relacionado: En las horas punta donde todos los autobuses están en uso o cuando la aerolínea sólo exige un autobús para todo el proceso de embarque, el tiempo necesario para completar el embarque es más largo; Por lo tanto, el embarque debe comenzar antes;
- Tiempo necesario para que el autobús lleve a los pasajeros desde el edificio de la terminal a la posición de estacionamiento correspondiente: Las aeronaves en posiciones remotas requieren más tiempo para embarcar porque la distancia entre la terminal y el estacionamiento puede ser larga y puede requerir más de un viaje en autobús para embarcar a todos los pasajeros.

Por ejemplo, en los puestos remotos, si el embarque comienza antes de lo que debería haber comenzado, los pasajeros esperan en frente del avión sin salir de los autobuses de embarque. Si un embarque en un parking remoto comienza tarde, aunque la limpieza y el abastecimiento de combustible terminen a tiempo, debido al hecho de que los autobuses llegan tarde al puesto de estacionamiento, el avión tendrá un retraso. La siguiente figura resume este tema.

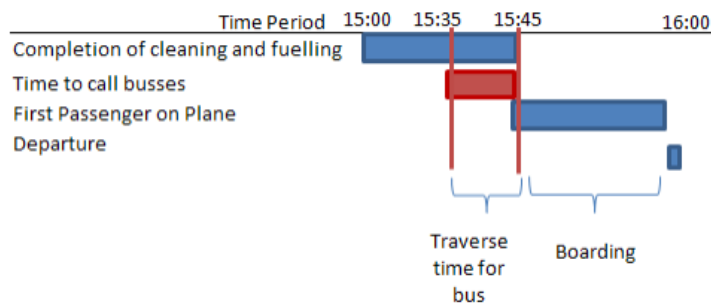


Figura 17 - Ejemplo de planificación en diagrama para embarque en parking remoto

Preparación y Aprobación de Carga y Balance

La Carga y Balance es un formulario que se rellena por la persona a cargo con la información recibida de los pilotos y la información de carga útil real. Con el fin de llenar la hoja de carga, la carga útil real tiene que ser conocida. La carga útil real sólo puede estar disponible después de que todos los pasajeros se registren. Después del 'check-in', el número exacto de los pasajeros y las bolsas se pasan a la persona que es responsable de preparar el formulario y el balance. Esta persona puede ser personal de operaciones o representante de una aerolínea. Además del equipaje y la información de los pasajeros, la cantidad de carga también se recibe del departamento de carga, y la cantidad de combustible de la aeronave se recibe para completar la hoja de carga. [10]

Este proceso incluye cambios de última hora debido a elevaciones de carga, pasajeros y combustible de última hora. Cuando la carga útil es más de lo esperado, la aerolínea decide retirar parte del peso por razones de seguridad. Por ejemplo, el representante de la aerolínea decide cargar o descargar carga, o por el contrario, el piloto puede decidir descargar el combustible para equilibrar el avión.

El personal de operaciones hace los ajustes finales en los cálculos en el balance y lo presenta al capitán para su aprobación. Sin la firma del capitán en esta hoja, el 'turnaround' de la aeronave no puede ser completado. Cuanto más rápido se haya completado la hoja de carga, más rápido podrá comenzar el 'pushback'. Si el capitán no está de acuerdo con los cálculos de carga y carga, puede ser necesario rellenarla de nuevo.

Figura 18 - Ejemplo hoja de carga. Boeing 757

Pushback

El proceso de 'Handling' de la aeronave se termina cuando se quitan los calzos -que se denomina 'chock-off time'- y se registra el tiempo real de salida anunciándose a todas las partes relacionadas a través de mensajes de movimiento.

El 'Push-back' está conectado a la aeronave, y después de la retirada de calzos, el 'push-back' comienza su proceso. El objetivo del retroceso es arrastrar el avión desde el puesto de estacionamiento en la dirección hacia atrás del avión hasta que el avión se gira en la dirección de la calle de rodaje, sin arrancar motores ni consumir combustible. Hay diferentes tipos de camiones de empuje: el que se puede llamar 'push-back' tradicional requiere las barras de remolque para cada tipo de avión, y los otros tipos son los nuevos sin usar una barra de remolque. Cerca del final del proceso de retorno, el auto de retroceso está conectado a la aeronave. Todo el equipo necesario en el proceso de empuje, como auriculares y clavijas, tiene que estar en el puesto de estacionamiento.

Los pilotos no pueden percibir obstáculos durante el retroceso, por lo que el personal de operaciones es el que puede ver a su alrededor para evitar los obstáculos mientras empuja hacia atrás el avión. El personal a cargo del empuje debe ser consciente de los movimientos de la aeronave y otros vehículos. Antes de comenzar a retroceder, los técnicos o el personal de operaciones caminan alrededor de la aeronave para limpiar el área de objetos extraños (FOD). Este control final se llama 'Walk-around-check' (recorrido de supervisión). Este proceso está diseñado para evitar cualquier accidente que se produzca debido a los obstáculos alrededor de la aeronave.

2.- ACDM: AIRPORT COLLABORATIVE DECISION MAKING

2.1.- ACDM

ACDM es el acrónimo de 'Airport Collaborative Decision Making'. Este es un concepto que tan sólo tiene sentido si se desarrolla en colaboración entre todos los actores intervinientes en el 'ground handling' aeroportuario de manera que se pueda mejorar la eficiencia del aeropuerto. La idea subyacente en este concepto consiste en compartir toda la información aeronáutica disponible con todos los agentes interesados en los procesos aeroportuarios (aeropuertos, líneas aéreas, proveedores de servicios de navegación aérea y agentes de 'ground handling') de manera que todos ellos tengan una perspectiva igual de la gestión del tráfico aéreo ('air traffic management', ATM) y las operaciones aeroportuarias. [2]

El ACDM enfatiza la necesidad de tomar decisiones de forma cooperativa generando así una conciencia común de la situación, y como consecuencia facilitando maximizar la eficiencia y por tanto la capacidad de los aeropuertos. Esto no sólo tiene como resultado una situación beneficiosa para todos los agentes intervinientes en el 'ground handling', sino que también es beneficiosa para los pasajeros ya que esto ayuda a hacer su viaje más sencillo. Estos agentes deben trabajar conjuntamente, compartiendo información en tiempo real, para así poder ser más eficientes y transparentes. El objetivo último es que la información apropiada le llegue al agente correspondiente con suficiente calidad y en el momento oportuno. Como consecuencia, debe existir un proceso de toma de decisiones mejor gracias a la información puntual y precisa. Las operaciones en tierra necesitan ser mejoradas para este fin, lo cual debe finalmente redundar en costes operacionales más bajos, menos ruido y emisiones de CO2 inferiores.

Debido a la propia naturaleza del ATM (se trata de una red), decisiones que parecen afectar únicamente de forma local pueden tener implicaciones en otros países.

La salida de un vuelo que se retrase en Barajas (Madrid) puede causar complicaciones en Gatwick (Londres), por ejemplo. Si el avión debe realizar en primer lugar el trayecto Barajas-Gatwick para después cubrir la ruta Gatwick-Malpensa (Milán), en el momento en que el primer vuelo se retrasa el segundo sufrirá probablemente retraso también, pudiendo llegar a ser cancelado en casos extremos. Ese es el principal motivo por el que el ACDM parece una herramienta espléndida para ayudar a solventar problemas que puedan surgir, mejorando la previsibilidad de eventos y el uso de los recursos disponibles.

Permite dar una respuesta rápida a cambios (retrasos, condiciones meteorológicas, etc). El resultado no es solo una mejora en la eficiencia localmente, sino una mejora de la eficiencia de la red como conjunto.

Es muy difícil desglosar las variables que intervienen en la llegada tardía de un vuelo. Estos minutos de retraso en la llegada, no siempre vienen dados por el mismo factor, ni mucho menos. Además en la mayoría de casos, se debe el retraso a diferentes factores. Es decir, quizás la llegada tardía de un vuelo, se deba a que ha salido con retraso, durante el vuelo ha tenido condiciones meteorológicas adversas, y además el tiempo de rodaje ha sido afectado por algún imprevisto. Es por eso, que se comprueba en la gran cuantía de variables que pueden interferir en los tiempos de tránsito de una aeronave.

De acuerdo con el manual de EUROCONTROL los objetivos de este método son:

- Mejorar la previsibilidad
- Mejorar la puntualidad
- Reducir los costes de los movimientos en tierra
- Optimizar el uso de los recursos de 'ground handling'
- Optimizar el uso de los puestos de estacionamiento, puertas de embarque y terminales
- Optimizar el uso de las infraestructuras aeroportuarias y reducir la congestión
- Reducir el desperdicio de slots aeroportuarios
- Permitir una planificación más flexible antes del vuelo
- Reducir la congestión en las plataformas y calles de rodaje

Para poder alcanzar todos estos objetivos no sólo hace falta un cambio de mentalidad con respecto a la manera en que se utiliza la información y se toman las decisiones, sino que también hace falta que los agentes posean ciertas habilidades y/o infraestructuras que les permitan hacer que el concepto del ACDM sea viable.

Hay ciertos conceptos que pueden ser utilizados con distinto significado a nivel internacional, o incluso a nivel de operador. Esto representa una barrera importante en el entendimiento entre las distintas partes implicadas, impidiendo que se produzca una comunicación fluida. Es por tanto de vital importancia darle un significado estandarizado a los conceptos para que la coordinación sea lo más sencilla posible y sin que se dé lugar a ninguna duda ni malentendido. EUROCONTROL ha sido el organismo encargado de establecer los cimientos para la implantación exitosa del ACDM gracias a la estandarización. El ACDM puede ser aplicado en distintos grados. No todos los aeropuertos necesitan aplicarlo por igual. Múnich y Bruselas fueron tomados como aeropuertos pilotos para la aplicación del método.

2.2.- ¿Por qué ACDM?

En la actualidad, muchos países se encuentran sumidos en una crisis económica global, una situación en la que presumiblemente todos los países se centran en intentar ahorrar o gastar lo mínimo posible. En el contexto aeroportuario esto puede llevarse a cabo mejorando la eficiencia, que es exactamente lo que el ACDM propone. Las mejoras necesarias se pueden lograr sin realizar inversiones prominentes de capital. [3]

Como se menciona anteriormente, el concepto del ACDM se implementa en el entorno aeroportuario por medio de la introducción de un conjunto de procedimientos operacionales y procesos automatizados. Ese es un buen motivo por el que el presente proyecto se desarrolla bajo el marco del ACDM, ya que mi principal intención es la de hacer las comunicaciones entre operadores de 'ground handling' y managers más automática, lo que debe redundar en una mejora en la eficiencia de las operaciones de 'ground handling' en los aeropuertos.



Figura 19 - Información compartida por los actores del ACDM

El concepto básico de comunicación en lo que concierne a los operadores de ‘ground handling’ en los aeropuertos se ilustra en la figura anterior. La idea es la de extender el concepto de gestión colaborativa a cada uno de los bloques que aparecen en la figura, y más específicamente pretendo realizar una simplificación introduciendo automatismos en el bloque del ‘ground handling’. Estos automatismos podrían ser extendidos al resto de comunicaciones, pero no es ese el objetivo en este caso.

La gestión del ‘ground handling’ no sólo contribuye en la implementación del ACDM, sino que también puede beneficiarse de ella. Como he mencionado con anterioridad, el ACDM pretende mejorar la previsibilidad, lo que en el caso particular del ‘ground handling’ podría significar que los managers fueran capaces de prever los recursos requeridos por un avión a su llegada y movilizarlos en el momento indicado. Gracias al ACDM, el ‘ground handling’ puede lograr una mejora en la puntualidad y una reducción en los costes operacionales.

2.3.- Beneficios del ACDM

Todos los beneficios tangibles atribuibles a A-CDM se realizan debido a la mejora en uno o más de los siguientes mecanismos de beneficios:

- Predictibilidad de llegada;
- Predictibilidad durante ‘turnaround’;
- Predictibilidad de despegue.

Como se ilustra en la siguiente figura, estos mecanismos están respaldados por facilitadores técnicos y procedimentales. Las mejoras en la predictibilidad del despegue son apenas posibles sin mejoras en la predictibilidad de ‘turnaround’, mientras que la previsibilidad de la llegada ayuda, pero no es crucial para las mejoras en la predictibilidad de ‘turnaround’.

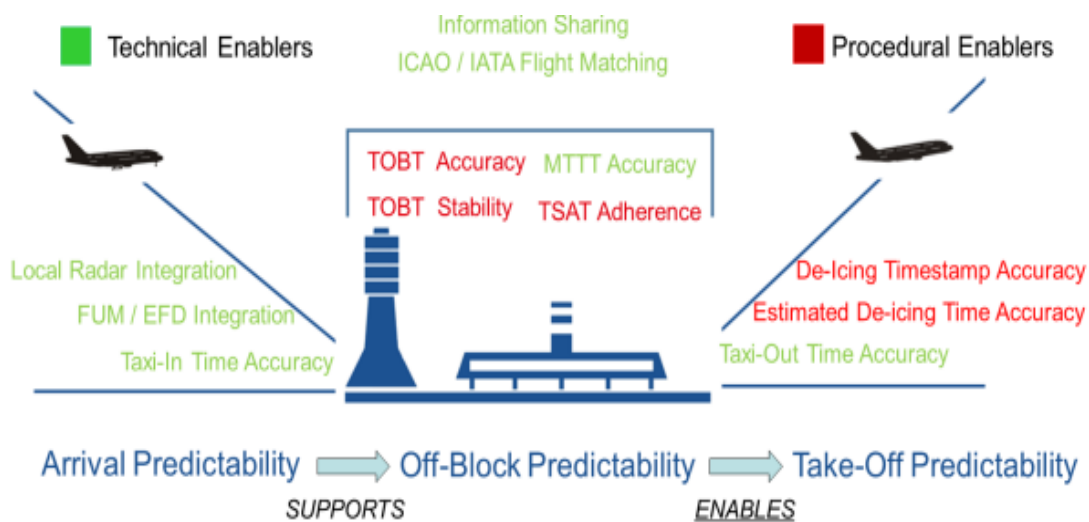


Figura 20 - Mecanismos de beneficio de A-CDM y ejemplos de facilitadores

Los facilitadores descritos en la última figura son sólo indicativos. Por ejemplo, la previsibilidad de la llegada puede ser apoyada por un procedimiento para refinar el tiempo de aterrizaje estimado (ELDT) basado en la posición de la aeronave en la secuencia de llegada o pila de espera. Algunos de los facilitadores "procedimentales" pueden implementarse sistemáticamente. Sin embargo, sin un cierto grado de intercambio de información y de adhesión procesal entre las partes interesadas, es evidente que nunca se pueden generar mejoras en la predictibilidad de 'turnaround' y en el despegue.

Los beneficios tangibles obtenidos de cada mecanismo de prestación dependen en gran medida de las características del aeropuerto y de la "oportunidad" de mejora en toda la operación. Actuar como un "filtro" - las limitaciones locales - puede suprimir todo el potencial de ganancias de desempeño donde otras prioridades operacionales tienen prioridad.

2.4.- Beneficios de Predictibilidad en la Llegada.

El A-CDM se centra en el principio de que un vuelo de salida es fundamentalmente una continuación y re-identificación de un vuelo de llegada que transita a través de una fase de "trayectoria de tierra". El recibo de mensajes de actualización de vuelo (FUM) proporciona un tiempo de aterrizaje estimado más preciso (ELDT) tan pronto como 3 horas desde el momento del aterrizaje. Esta información proporciona a los agentes del aeropuerto la información que necesitan para asignar mejor los recursos si el ELDT de un vuelo de llegada cambia significativamente.

FUM se distribuyen a través del servicio web EUROCONTROL B2B o de la Red de Telecomunicaciones Fija ATFM. La distribución de actualizaciones de llegada fiables entre los socios del aeropuerto ha demostrado mejoras en las siguientes áreas:

- Planificación de los stands
- Asignación de recursos de 'Handling'
- Planificación de la flota
- Puntualidad de salida

Planificación de stands

El tiempo estimado de puesta de calzos (EIBT) se genera automáticamente a partir del ELDT y se calcula el tiempo de rodaje (EXIT) y se ha demostrado que soporta una mejor utilización de los stands. Resultando en menos casos, congestión de los stands o ‘turnarounds’ tardíos. Este beneficio se aplica especialmente a los aeropuertos donde la capacidad de los puestos tienen limitaciones o donde el tráfico restringe el uso total de todos los activos del stand durante los picos operacionales.

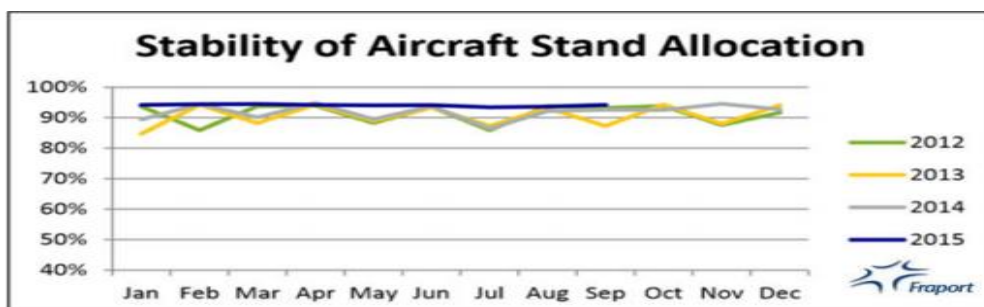


Figura 21 - Estabilidad de la utilización de stands en el aeropuerto de Frankfurt en 2015

Asignación y priorización de los recursos de ‘Handling’

Los operadores de tierra (‘Ground Handling’) están sujetos a acuerdos de nivel de servicio (SLA) para cumplir con los objetivos de puntualidad de salida, a menudo con fuertes incentivos financieros. El desempeño en tiempo (‘On Time Performance’) es su prioridad, pero esto debe lograrse con recursos limitados. La previsibilidad del tiempo de llegada está ayudando a los Operadores de ‘Handling’ a planear su operación basándose en la evolución de la información táctica en lugar de los calendarios, en gran parte imprecisos.

- Durante el retraso severo del ‘Ground Handling’, la previsibilidad del tiempo de llegada permite la priorización de los vuelos y un diálogo mejorado entre el Centro de Control Operacional de la Línea Aérea (OCC) y los representantes del ‘Handling’.
- Hay menos casos en los que los vuelos no se cumplen en tiempos del personal que espera hacer el ‘turnaround’ de la aeronave. Junto con una mayor robustez de la asignación de los puestos, la previsibilidad mejorada de llegada permite posicionar los recursos en el puesto correcto en el momento adecuado.
- El tiempo de inactividad del recurso se reduce y se aumenta la utilización. Si un vuelo entrante se retrasa, entonces esto representa una oportunidad para que la GH reasigne recursos para proteger a la OTP de otros vuelos.

La evidencia que se ha recopilado para apoyar los beneficios de la previsibilidad de llegada en la función de manejo en tierra es totalmente cualitativa. El proyecto no ha sido lo suficientemente cercano a las agencias de ‘Handling’ para generar resultados cuantitativos firmes. Sin embargo, la evidencia anecdótica y las cuentas personales apoyan que la información de ELDT presentada a los manejadores de tierra tan pronto como 3 horas por adelantado tiene ventajas directas en sus propias operaciones.

Planificación de la flota

El OCC de la línea aérea también se beneficia de la predictibilidad mejorada del tiempo de llegada. Un avión que está planeado para volar varios sectores en el transcurso del día puede ser re-planteado (o cancelado) proactivamente basado en las notificaciones de retraso recibidas durante las secciones anteriores.

La siguiente figura ilustra cómo la previsibilidad del tiempo de llegada de un vuelo es afectada por el A-CDM tanto en los aeropuertos de partida como de destino. En el ejemplo siguiente, London Heathrow (LHR) y Paris Charles De Gaulle (CDG) representan respectivamente el aeropuerto de salida y el aeropuerto de destino. Los números anotados, representan diferentes fases de la trayectoria durante las cuales la precisión en el tiempo del vuelo de retorno podría mejorar.



Figura 22 - Impacto del A-CDM en la previsibilidad del tiempo de llegada

Reacción a retrasos y rendimiento del Turnaround

La salida tardía del vuelo es causada por uno o más de los siguientes factores:

1. Llegada tardía del vuelo entrante.
2. No hacer el 'turnaround' (fase de trayectoria en tierra) dentro del tiempo requerido.
3. Retraso ATFM (CTOT) asignado a la salida.
4. Las limitaciones locales de ATC o infraestructura aeroportuaria que pueden impedir una puesta en marcha inmediata y el despeje de taxi.

Los manejadores en tierra se benefician de mejoras en la previsibilidad de llegada y salida. No se sabe hasta qué punto el desempeño en algunos de los aeropuertos del CDM están influenciados por el componente de mejora de previsibilidad de la llegada.

2.5.- Principales elementos del A-CDM

El 'Target Off-Block Time' (TOBT) y el 'Target Start-Up Approval Time' (TSAT) son los elementos de datos más importantes dentro del proceso ACDM.

El TOBT se define como el momento en que el operador de la aeronave o el operador de tierra estiman que una aeronave estará lista, todas las puertas cerradas, el 'finger' de embarque retirado, el 'Push-back' disponible y listo para arrancar / empujar inmediatamente después de la recepción del ATC. El TOBT debe ser exacto a 5 minutos del actual off-block time (AOBT). El procedimiento TSAT es un mecanismo de planificación transparente y flexible antes de la partida del avión.

2.5.1.- Procedimiento TSAT

El procedimiento TSAT es un mecanismo de planificación transparente y flexible antes de la partida del avión. El TSAT es propiedad del ATC y generalmente es generado por un secuenciador antes de la salida (PDS) o administrador de salida (DMAN). El TSAT es el tiempo que se espera que el ATC despeje el avión para arrancar y empujar el motor. El TSAT nunca puede ser anterior al TOBT y debe tener en cuenta las restricciones locales de ATC y de infraestructura aeroportuaria tales como la congestión del suelo, la contención de la pista, la demanda de la pista y los 'SLOTS' de ATFM.

El TSAT refleja el equilibrio de la infraestructura y la capacidad del espacio aéreo con la demanda generada por el TOBT.

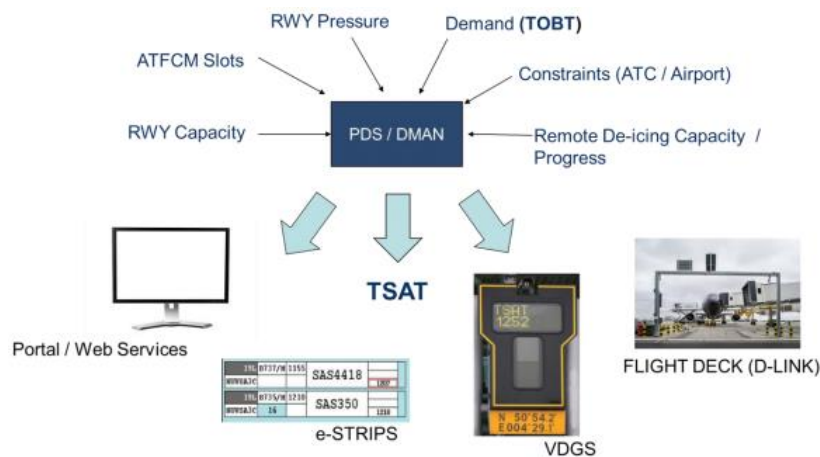


Figura 23 - Factores que afectan al valor TSAT y ejemplos de métodos de publicación

Las mejoras en la predictibilidad fuera de calzos son impulsadas por la estabilidad del TOBT, la estabilidad de la secuencia y la adherencia del TSAT. Si las actualizaciones de TOBT son tardías o inexactas, entonces no es razonable esperar una secuencia óptima o estable previa a la salida. También es vital que el ATC pueda facilitar una autorización de inicio cercana al TSAT asignado.

El procedimiento TSAT introduce las siguientes ventajas operacionales a un aeropuerto CDM:

1. Un punto de medición para medir y supervisar la aprobación de arranque y el retraso de 'Push-back'.
2. Un medio transparente y equitativo para absorber el retraso del taxi en el stand ('Green delay').

3. Optimización de la secuencia de salida de la pista para maximizar su rendimiento (asumiendo que un DMAN no estaba disponible antes de la implementación del A-CDM).
4. Una referencia para mejorar la planificación de recursos y activos que es más fiable que el plan de vuelo EOBT o los horarios de los aeropuertos.
5. Una secuencia previa a la partida consistente en vuelos que exigirán una autorización de puesta en marcha más previsible.
6. Reducción de la Posición de Gestión de Flujo ('Flow Management Position') en las regulaciones de salida emitidas. Los caudales aplicados se pueden gestionar directamente dentro de la secuencia previa a la salida. [8]

La conciencia de la situación común entre todas las partes interesadas puede tener un coste para el operador del aeropuerto. La transparencia de las marcas de tiempo operativas puede conducir a que las partes interesadas formulen sus propias conclusiones sobre la equidad del retraso del TSAT. Para manejar mejor este riesgo, las comunicaciones del equipo del proyecto A-CDM pueden ser necesarias para convencer a las partes interesadas de que:

1. No se favorece ninguna única operadora en la generación de la secuencia previa a la partida.
2. Que el retraso TSAT en el 'turnaround' no retrasa el tiempo real de despegue.
3. Que el retraso de TSAT ha dado como resultado la reducción de tendencia en los tiempos medios de rodaje-salida para todos (si corresponde).

El proceso de puesta en marcha de los aeropuertos A-CDM es más eficiente que el caso previo a la implementación. En todos los casos en los que fue posible evaluar, la adherencia TSAT (la proporción de vuelos con una aprobación inicial de TSAT +/- 5') ha mostrado un aumento de tendencia de tan poco como 40%. La siguiente figura ilustra la mejora en la adherencia a TSAT en Londres Heathrow desde la implementación y es típica de la mayoría de los aeropuertos de CDM para los cuales los datos se pusieron a disposición para el análisis.

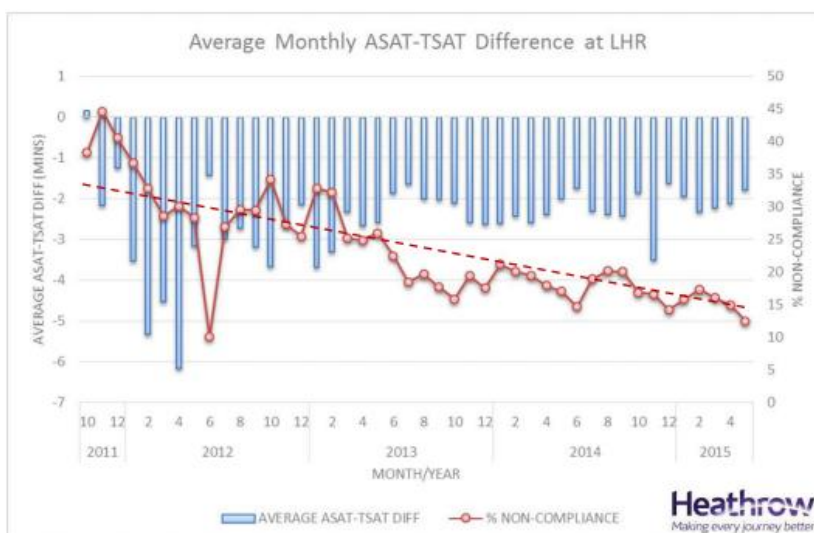


Figura 24 - Tendencia de cumplimiento de TSAT en Londres Heathrow desde octubre 2012 hasta el 2015.

El papel del ATC en asegurar la adherencia del TSAT. Es el último eslabón crítico en la cadena de proceso. El proceso del hito A-CDM está diseñado para proporcionar operaciones de puesta en marcha predecibles para impulsar la previsibilidad de despegue requerida por los sectores ATC posteriores. Los factores que impulsan la mejora en la adherencia TSAT incluyen:

- Aumento del cumplimiento de la secuencia anterior a la salida, como sugiere el DMAN / PDS.
- Aumento de la adopción de solicitud / autorización de inicio sobre enlace de datos.
- Estabilidad mejorada de la secuencia de salida. Es difícil para el ATC facilitar una secuencia de salida estable, especialmente en las situaciones de congestión. La estabilidad de la secuencia de salida se promueve por el aumento de la estabilidad del TOBT y la previsión de demanda de la pista.

Tal vez la mayor barrera para lograr la adherencia consistente del TSAT es el nivel de confianza en los sistemas PDS y / o DMAN, especialmente en grandes situaciones de retraso en las que el retraso del TSAT puede causar más carga de trabajo.

2.5.2.- TOBT UPDATE

El 'TOBT Updater' está en el corazón mismo del proceso A-CDM. Los altos niveles de predictibilidad local y de la red no pueden ser alcanzadas a menos que una cultura de actualizaciones tempranas y precisas de TOBT esté incrustada en el aeropuerto del CDM. Promover y mantener un buen comportamiento de actualización del TOBT es un esfuerzo constante para el equipo de proyecto A-CDM.

Sin embargo, debe reconocerse que el Operador de tierra está sirviendo al Operador de Aeronaves quien a veces prefiere que no se realice una actualización de TOBT cuando se muestra un retraso, especialmente cuando el vuelo tiene una regulación. Esto pone a los operadores de tierra en una posición poco envidiable donde no pueden adherirse completamente a los procedimientos de A-CDM debido a una "cultura de culpa" que resulta de cualquier retraso local o de ATFM que puede resultar en una entrada de TOBT más precisa.

Los Operadores de tierra pueden ser reacios a proporcionar actualizaciones tempranas de TOBT porque:

1. Un vuelo que no está actualmente sujeto a un reglamento (basado en el EOBT) podría quedar sujeto al retraso ATFM después de la actualización TOBT.
2. Se puede incurrir en una demora adicional del TSAT por mover el vuelo en un período de mayor demanda de salida.
3. El vuelo se encuentra dentro de la zona de secuenciación y el operador no quiere ser responsable de que el vuelo sea reordenado en la secuencia debido a la superación del límite de actualización de TOBT.
4. Los operadores reciben instrucciones explícitas de no actualizar el TOBT en ciertas situaciones de retraso. Por ejemplo, el OCC puede tener la flexibilidad de realizar un swap de cola y no desea que una regulación aplicada se modifique al cambiar una aeronave, que mantendrá el vuelo en su plan de vuelo original.

Se acepta que las actualizaciones del TOBT puedan darse en un vuelo que está siendo sometido a una regulación, sin embargo, cuanto antes se comunique el retraso a NMOC mejor será la disponibilidad de 'SLOT' dentro del flujo de tráfico restringido. Si el retraso es inevitable y no se recupera, entonces una actualización TTOT a NMOC (a través del DPI) sirve para mejorar las posibilidades de que el vuelo reciba el mejor 'SLOT' posible - antes de que sean ocupados por otros vuelos. También debe mencionarse que una actualización TOBT podría darse en un vuelo regulado que se lleva fuera de la regulación y por lo tanto su CTOT se cancela.

El TOBT permite optimizar la secuencia de salidas de aeronaves en un aeropuerto, ya que al actualizar el tiempo de TOBT, se pueden proporcionar datos para el ATM y que estos puedan organizar su salida.

En los aeropuertos que no disponen de CDM, y por lo tanto no pueden trabajar con el TOBT, el sistema que se suele utilizar es el 'First Come First Serve' (FCFS), donde la salida de los vuelos es secuenciada según su llegada. Este tipo de gestión es muy poco habitual en los aeropuertos con un flujo de pasajeros alto. Este sistema, concede la salida según la llegada de la aeronave. Su significado es que la primera aeronave en llegar, será la primera aeronave en salir. Esto provoca una peor utilización de la pista de despegue, ya que con el TOBT y una secuencia de salidas según el TOBT de la aeronave, se aprovecha mejor el tiempo entre salida y salida de vuelo.

Actualizaciones tempranas de TOBT y comprobación de la lista de aeronaves

En algunos aeropuertos del CDM, algunas aerolíneas siguen aplicando las actualizaciones TOBT "tempranas" o demasiado optimistas para forzar una mejor posición en la secuencia de salida. Esta práctica demuestra que algunas partes interesadas aún no están suscritas al elemento "colaborativo" del proceso y están dispuestas a comprometer la previsibilidad de una operación global para su propio beneficio a corto plazo. Los procesos de A-CDM están diseñados para bloquear y disuadir dicho comportamiento a través de medidas tales como:

- Inspección de aeronaves 'Aircraft Ready Check' - realizada por los aeropuertos italianos del MDL antes de la solicitud de puesta en marcha.
- Puesta en marcha del procedimiento de TOBT. En la medida de lo posible, los aeropuertos implementan actualmente un control preparado cuando la tripulación solicita la puesta en marcha dentro del rango $TOBT + / 5'$.
- Adhesión TSAT: los vuelos que no están realmente listos dentro de la ventana TSAT están suspendidos y deben proporcionar un TOBT nuevo (que es 5' mayor) antes de re-secuenciarse.

La verificación de adherencia TSAT es la última línea de defensa para prevenir entradas de TOBT imprecisas que resultan en una autorización de inicio exitosa. Sin embargo, para los aeropuertos que implementan un procedimiento de "Call on TSAT" sin verificación "lista" durante o antes del TOBT, este procedimiento todavía está muy abierto al abuso.

Los vuelos que declaran un TOBT que está antes del TOBT real lo harán para recibir un TSAT favorable en situaciones de retraso alto. Si este TSAT es mayor que la actualización real de TOBT del vuelo, entonces el vuelo habrá logrado explotar el proceso si no hay un control preparado en el aeropuerto. Sin embargo, si el TSAT es anterior al TOBT real, el juego dará lugar a un retraso adicional para todos. [8]

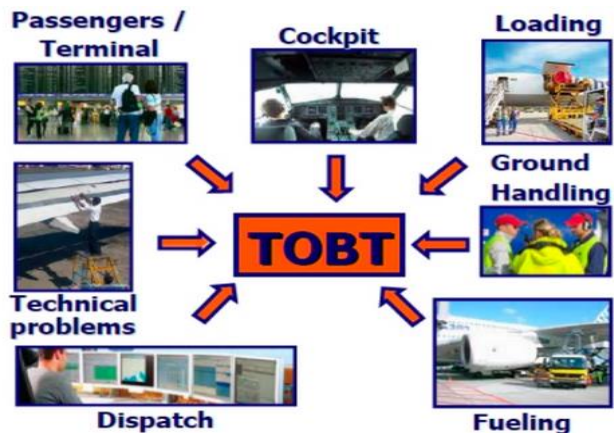


Figura 25 - Representación visual de participantes en el TOBT

3. ANALISIS DE LOS MODELOS PREDICTIVOS

La operativa aeronáutica en general, requiere de constante supervisión y control. Una operativa constantemente afectada por multitud de factores externos e internos al propio operador, necesita de unos habituales y reiterados procedimientos de análisis de datos y modelos predictivos, con el fin de evitar y/o apaciguar los efectos nocivos de las inferencias en la operativa normal, así como con el fin de prever comportamientos y optimizarlos para lograr aumentar la eficacia y eficiencia de las acciones.

Los modelos predictivos analizan los resultados anteriores para evaluar qué probabilidad tiene un cliente para mostrar un comportamiento específico en el futuro con el fin de mejorar la eficacia de una operación.

Una teoría científica cuyas predicciones no son corroboradas por las observaciones, por las pruebas o por experimentos probablemente será rechazada. Las teorías que generan muchas predicciones que resultan de gran valor (tanto por su interés científico como por sus aplicaciones) se confirman o se falsean fácilmente y, en muchos campos científicos, las más deseables son aquellas que, con número bajo de principios básicos, predicen un gran número de sucesos.

El análisis predictivo utiliza la estadística junto con algoritmos de 'data mining'. Se basan en el análisis de los datos actuales e históricos para hacer predicciones sobre futuros eventos. Dichas predicciones raramente suelen ser afirmaciones absolutas, pareciéndose más a eventos y su probabilidad de que suceda en el futuro.

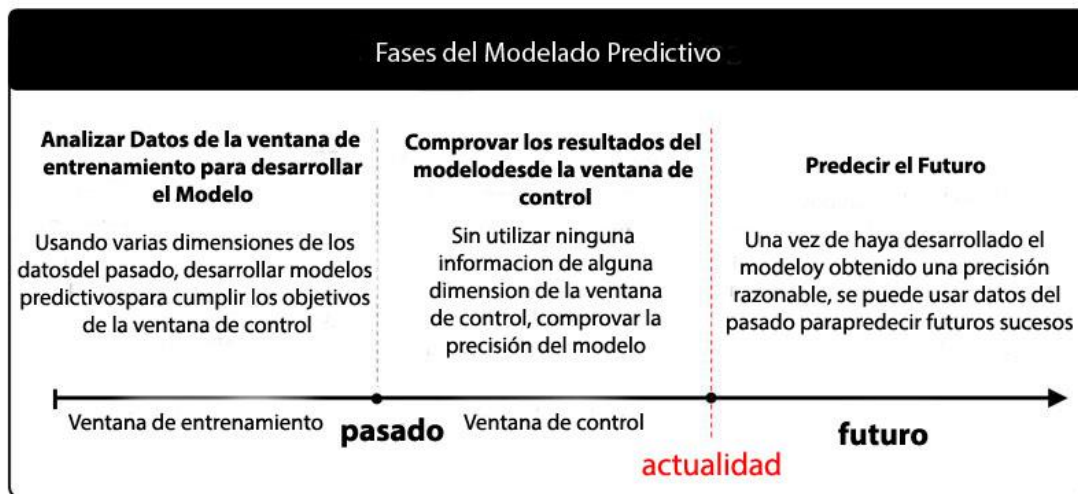


Figura 26 - Fases del modelado predictivo

El modelo de predicción utiliza modelos estadísticos para la producción de outputs. Habitualmente se pretenden predecir acciones o eventos que sucederán en un futuro, lejano o inmediato, pero los modelos de predicción se pueden aplicar a cualquier tipo de evento desconocido, independientemente del momento en que ocurrió. Por ejemplo, los modelos predictivos se utilizan a menudo para detectar delitos e identificar a los sospechosos, después de que un crimen haya tenido lugar.

En muchos casos se elige el modelo sobre la base de la teoría de detección para tratar de adivinar la probabilidad de un resultado dado una cantidad de datos de entrada, por ejemplo, dado un correo electrónico se intenta determinar qué tan probable es que se trate de un correo spam.

En los modelos se pueden utilizar uno o más clasificadores para tratar de determinar la probabilidad de un conjunto de datos que pertenecen a otro grupo, es decir spam o 'publicitario'.

En función de los límites de definición, modelado predictivo es sinónimo de, o en gran medida con el ámbito del aprendizaje de máquina, como se le conoce más comúnmente en contextos académicos o de investigación y desarrollo. Cuando se implementa en el comercio, el modelado predictivo se conoce como análisis predictivos a menudo.

En el caso de la predicción, por ejemplo, es necesario predecir si cambiará la dirección del viento para poder optimizar el sistema de tráfico aéreo y elegir la pista adecuada para que las aeronaves aterricen o despeguen con viento en cara, o re-enrutar su aproximación en base a este cambio de dirección. Una predicción fiable y de precisión maximiza la previsión de ocurrencia de ciertos fenómenos vitales en la actividad aeronáutica.

Las predicciones tienen una gran importancia en el sector aeronáutico, como por ejemplo, la predicción de tiempos de llegada de una aeronave. Para dicha predicción, necesitaremos un análisis de datos como inputs (Información meteorológica, tiempos de entrada y salida en el aeropuerto de origen, tiempo medio en espera, o la velocidad media de la aeronave) con el propósito de obtener unos outputs cuyo resultado nos será de gran ayuda en la toma de decisiones y en la creación de afirmaciones predictivas.

Otro claro ejemplo sería la predicción del punto álgido de movimiento de pasajeros en un aeropuerto en temporadas críticas, como la época estival. Para esta predicción necesitaremos otro tipo de datos como inputs, como el volumen de pasajeros de los años anteriores, la situación de la economía actual, el número de reservas o el porcentaje de ocupación hotelera. Con el análisis de dichos datos se obtendrán los resultados que nos proporcionarán información vital para el desarrollo de predicciones fiables.

Dichas predicciones forman una de las partes más importantes de la planificación de operaciones aeronáuticas. Una predicción fiable permite crear modelos de planificación más precisos y eficaces, ahorrando una considerable cantidad de tiempo y recursos a todas las empresas e instituciones que forman parte de la operativa diaria aeronáutica.

Una buena predicción en la llegada de una aeronave, como en el ejemplo anterior, permite organizar las tareas de embarque y escala de forma precisa y continuada, evitando los 'standby moments' y reduciendo tiempo y recursos. Así, una buena predicción también en el volumen de pasajeros en un intervalo de tiempo específico, permite asignar los recursos justos y necesarios a dicha demanda, prever los colapsos y congestiones y mejorar la atención al cliente.

Las predicciones permiten en general, mejorar la eficacia y la eficiencia de los sistemas actuales, crear nuevas operativas o modificaciones adaptadas a dichas predicciones y optimizar los procesos aeronáuticos, mejorando en tiempo, recursos e incluso en seguridad. [16]

3.1.- Modelo predictivo 'Random Forest'

'Random forest' es una combinación de árboles predictores tal que cada árbol depende de los valores de un vector aleatorio probado independientemente y con la misma distribución para cada uno de estos. Es una modificación sustancial de 'bagging' que construye una larga colección de árboles no correlacionados y luego los promedia.

En muchos problemas el rendimiento del algoritmo 'random forest' es muy similar a la del 'boosting', y es más simple de entrenar y ajustar. Como consecuencia el 'random forests' es popular y es ampliamente utilizado.

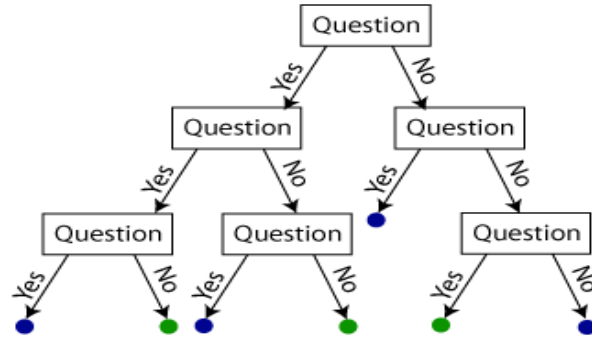


Figura 27 - Árbol de Random Forests

La idea esencial del 'bagging' es promediar muchos modelos ruidosos pero aproximadamente imparciales, y por tanto reducir la variación. Los árboles son los candidatos ideales para el 'bagging', dado que ellos pueden registrar estructuras de interacción compleja en los datos, y si crecen suficientemente profundo, tienen relativamente baja parcialidad. Producto de que los árboles son notoriamente ruidosos, ellos se benefician grandemente al promediar.

Cada árbol es construido usando el siguiente algoritmo:

1. Sea N el número de casos de prueba, M es el número de variables en el clasificador.
2. Sea m el número de variables de entrada a ser usado para determinar la decisión en un nodo dado; m debe ser mucho menor que M
3. Elegir un conjunto de entrenamiento para este árbol y usar el resto de los casos de prueba para estimar el error.
4. Para cada nodo del árbol, elegir aleatoriamente m variables en las cuales basar la decisión. Calcular la mejor partición a partir de las m variables del conjunto de entrenamiento.

Para la predicción un nuevo caso es empujado hacia abajo por el árbol. Luego se le asigna la etiqueta del nodo terminal donde termina. Este proceso es iterado por todos los árboles en el ensamblado, y la etiqueta que obtenga la mayor cantidad de incidencias es reportada como la predicción.

Características (o rasgos) y Ventajas

Las ventajas del 'random forests' son:

- Es uno de los algoritmos de aprendizaje más certeros que hay disponible. Para un set de datos lo suficientemente grande produce un clasificador muy certero.
- Corre eficientemente en bases de datos grandes.
- Puede manejar cientos de variables de entrada sin excluir ninguna.
- Da estimados de qué variables son importantes en la clasificación.
- Tiene un método eficaz para estimar datos perdidos y mantener la exactitud cuando una gran proporción de los datos está perdida.

- Computa los prototipos que dan información sobre la relación entre las variables y la clasificación.
- Computa las proximidades entre los pares de casos que pueden usarse en los grupos, localizando valores atípicos, o (ascendiendo) dando vistas interesantes de los datos.
- Ofrece un método experimental para detectar las interacciones de las variables.

Desventajas de 'Random Forest'

- Se ha observado que el 'Random forests' se sobre ajusta en ciertos grupos de datos con tareas de clasificación/regresión ruidosas.
- A diferencia de los árboles de decisión, la clasificación hecha por 'random forest' es difícil de interpretar por el hombre.
- Para los datos que incluyen variables categóricas con diferente número de niveles, el 'random forest' se parcializa a favor de esos atributos con más niveles. Por consiguiente, la posición que marca la variable no es fiable para este tipo de datos. Métodos como las permutaciones parciales se han usado para resolver el problema.
- Si los datos contienen grupos de atributos correlacionados con similar relevancia para el rendimiento, entonces los grupos más pequeños están favorecidos sobre los grupos más grandes.

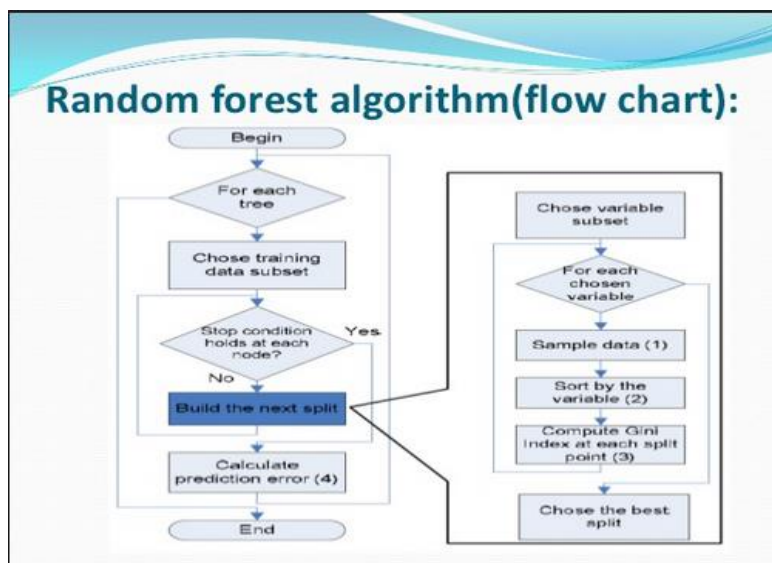


Figura 28 - 'Flow chart' del Random Forest

3.2.- Modelo predictivo 'Gradient Boosting Machine'

'Gradient Boosting' es una técnica de Aprendizaje Automático ('Machine Learning') para la regresión y clasificación de problemas, que crea un modelo predictivo en forma de conjunto de modelos débiles de predicción. Típicamente árboles de decisión. Esta técnica construye el modelo de una manera escalonada en etapas, al igual que otros modelos estilo 'Boosting', y los generaliza al permitir la optimización de una función de pérdida diferenciable arbitraria.

La idea de la técnica 'Gradient Boosting', se originó por la observación de Leo Breiman, donde el 'Boosting' puede ser interpretado como un algoritmo de optimización en una función de coste adecuada. Los 'Gradient Boosting' de regresión explícita, fueron desarrollados posteriormente, simultáneamente con el 'Gradient Boosting' más general y funcional. Los dos últimos trabajos introdujeron la visión abstracta de impulsar algoritmos como algoritmos de 'Functional Gradient Descent' (Descenso Gradual Funcional). Es decir, los algoritmos que optimizan un coste funcional sobre el espacio funcional por la elección de una función iterativa (Hipótesis Débil) que apunta en la dirección del gradiente negativo.

Este punto de vista funcional del 'Boosting' lo ha llevado al desarrollo, a impulsar algoritmos en muchas áreas de aprendizaje automático y estadísticas más allá de la regresión y la clasificación.

En conclusión, el método 'Gradient Boosting' representa un algoritmo de aprendizaje automático aplicable en ámbitos muy generales, y con el que se puede obtener un gran rendimiento.

Los modelos de impulso o 'boosting' fueron desarrollados originalmente para los problemas de clasificación y se extendieron posteriormente a la configuración de regresión. Esta historia comienza con el algoritmo 'AdaBoost' y evoluciona a la 'gradient boosting machine' estocástica de Friedman, que ahora es ampliamente aceptado como el algoritmo de 'boosting' de mayor elección, en cuyo caso aplicaremos en esta sección.

A principios de 1990 aparecieron los primeros algoritmos de 'boosting' "Schapire 1990", "Freund 199" y "Schapire 1999", que fueron influenciados por la teoría del aprendizaje "Valiant 1984" y "Kearns y Valiant 1989", en el que una serie de débiles clasificadores (un clasificador que predice un poco mejor que al azar) se combinan (o impulsan) para producir un clasificador conjunto con una tasa de error generalizada superior.

Los investigadores se esforzaron por un tiempo para hallar una implementación reflexiva de impulsar la teoría, hasta que Freund y Schapire colaboraron para producir el algoritmo 'AdaBoost' (Schapire 1999). 'AdaBoost' proporciona una implementación práctica del concepto de 'boosting'. Impulso o 'boosting', especialmente en la forma del algoritmo 'AdaBoost', ha demostrado ser una herramienta de predicción potente, por lo general superando cualquier modelo individual. Su éxito llamó la atención de la comunidad de modelado y su uso se generalizó con aplicaciones en la expresión génica quimiometría, y el género de la música. [15]

El algoritmo 'AdaBoost' claramente funcionó, y después de su llegada con éxito, varios investigadores (Friedman et al. 2000) conecta el algoritmo 'AdaBoost' a los conceptos estadísticos de funciones de pérdida, modelado aditivo y de regresión logística y mostró que el aumento se puede interpretar como un aditivo por etapas hacia adelante en un modelo que minimiza la pérdida exponencial. Esta comprensión fundamental de 'boosting' llevado a una nueva visión de impulso que facilitó a varias generalizaciones algorítmicas los problemas de clasificación.

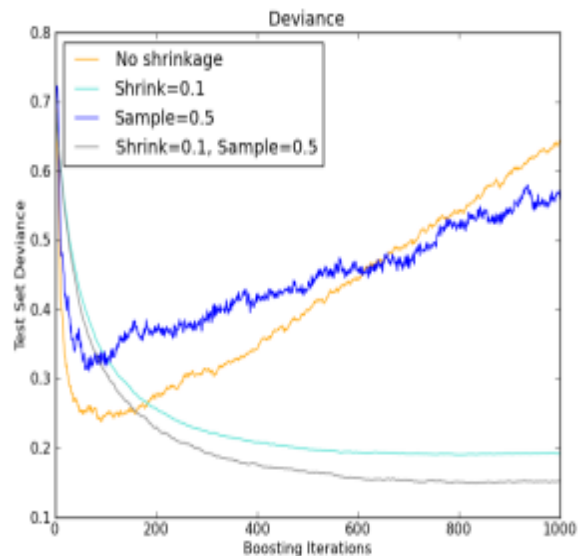


Figura 29 - Gradient Boosting Machine

Por otra parte, esta nueva perspectiva también permitió el método se extienda a problemas de regresión.

La capacidad de Friedman para ver el marco estadístico de 'boosting arrojó' un algoritmo simple, elegante y altamente adaptable para diferentes clases de problemas (Friedman 2001). Él llama a este método 'gradient boosting machine', que abarcaba tanto la clasificación como la regresión.

Los principios básicos de 'gradient boosting' son los siguientes: dado una función de pérdida (por ejemplo, error cuadrático para la regresión) y un principiante débil (por ejemplo, árboles de regresión), el algoritmo busca para hallar un modelo aditivo que minimiza la función de pérdida. El algoritmo se inicializa típicamente con la mejor estimación de la respuesta (por ejemplo, la media de la respuesta en la regresión).

- 1 Selecciona la profundidad del árbol, D , y el número de iteraciones, K
- 2 Computa la average response, y , y se usa como valor inicial predictor para cada muestra
- 3 **For** $k = 1$ to K **do**
- 4 Computa el residual, la diferencia entre el valor observado y el valor actual predicho, para cada muestra
- 5 Encaja un árbol de regresión de profundidad, D , usando los residuales como response
- 6 Predice cada muestra usando el árbol de regresión del paso anterior
- 7 Actualiza el valor de la predicción de cada muestra añadiendo el valor de la iteración anterior al valor de la predicción generado en el paso anterior
- 8 **end**

Figura 30 - Metodología GBM

‘Shrinkage’: Una parte importante del método de ‘gradient boosting’ machine es la regularización por la contracción o ‘shrinkage’, que consiste en la modificación de la regla de actualización de la siguiente manera:

$$F_m(x) = F_{m-1}(x) + \nu \cdot \gamma_m h_m(x), \quad 0 < \nu \leq 1,$$

Donde el parámetro ν es llamada la " tasa de aprendizaje".

Empíricamente se ha encontrado que el uso de pequeñas tasas de aprendizaje (como $\nu < 0,1$) produce mejoras en la capacidad de generalización del modelo de ‘gradient boosting’ machine sin encogerse ($\nu = 1$) Sin embargo, se trata al precio de aumentar el tiempo computacional, tanto durante el entrenamiento y consulta: a menor tasa de aprendizaje, más iteraciones.

Cantidad de categorías

Hay que tener en cuenta, y aunque no sea de aplicación en este proyecto, que para el procedimiento de la aplicación del ‘gradient Boosting’, en la clasificación de problemas, requiere secuencias separadas de los arboles (‘Boosted Trees’), cada una creada para cada categoría o clase.

Por lo tanto, el esfuerzo computacional en general se hace más grande, es decir, se multiplica, más de lo que es necesario para resolver un problema de predicción de regresión simple (para una sola variable dependiente continua). Por lo tanto no sería acertado este método para analizar las variables dependientes categóricas (variables de clase) con más de, aproximadamente, 100 o más clases. Pasado este máximo, los cálculos realizados pueden necesitar una cantidad razonable de tiempo y esfuerzo.

[17]

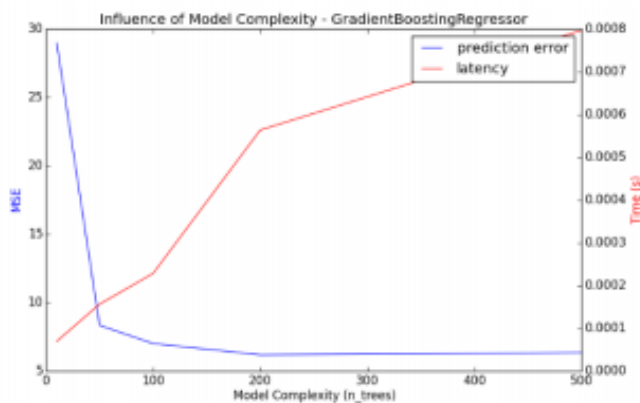


Figura 31 - Demostración cómo la complejidad del modelo influye tanto en la precisión de la predicción y el rendimiento computacional.

3.3.- Evaluación del modelo usando RMSE (Root-Mean-Square Error)

Con los siguientes modelos predictivos se obtendrán los resultados en forma de RMSE ('Root-Mean-Square-Error') o lo que es lo mismo la Raíz del Error Cuadrático Medio, pudiendo ser a la vez RMSD ('Root-Mean-Square Deviation') o lo que es lo mismo la Raíz de Desviación Cuadrática Media.

El Error Cuadrático Medio ('MSE- Mean Square Error') en estadística se trata de un estimador que mide el promedio de los errores al cuadrado, es decir, la diferencia entre el estimador y lo que se estima. El MSE es una función de riesgo, correspondiente al valor esperado de la pérdida del error al cuadrado o pérdida cuadrática. La diferencia se produce debido a la aleatoriedad o porque el estimador no tiene en cuenta la información que podría producir una estimación más precisa.

El MSE ('Mean Square Error') es el segundo momento (sobre el origen) del error, y por lo tanto incorpora tanto la varianza del estimador así como su sesgo. Para un estimador insesgado el RMSE es la varianza del estimador. Al igual que la varianza, el RMSE tiene las mismas unidades de medida que el cuadrado de la cantidad que se estima.

En una analogía con la desviación estándar tomando la raíz cuadrada del RMSE produce el error de la raíz cuadrada de la media o la desviación de la raíz cuadrada media (RMSE o RMSD), que tiene las mismas unidades que la cantidad que se estima, para un estimador insesgado, el RMSE es la raíz cuadrada de la varianza, conocida como la desviación estándar.

Fórmula

El RMSD de un estimador con respecto a un estimador parámetro es definido como la raíz cuadrada del error cuadrático medio:

$$\text{RMSD}(\hat{\theta}) = \sqrt{\text{MSE}(\hat{\theta})} = \sqrt{E((\hat{\theta} - \theta)^2)}.$$

Para un estimador insesgado, el RMSD es la raíz cuadrada de la varianza, conocido como el error estándar.

El RMSD de los valores previstos \hat{y}_t para tiempos t y de variable dependiente de una regresión se calcula para n diferentes predicciones como la raíz cuadrada de la media de los cuadrados de las desviaciones:

$$\text{RMSD} = \sqrt{\frac{\sum_{t=1}^n (\hat{y}_t - y)^2}{n}}.$$

En algunas disciplinas, el RMSD se utiliza para comparar las diferencias entre dos cosas que pueden variar, ninguno de los cuales es aceptado como el "estándar". Por ejemplo, al medir la diferencia media entre dos series de tiempo x_1, t y x_2, t , la fórmula se convierte en:

$$\text{RMSD} = \sqrt{\frac{\sum_{t=1}^n (x_{1,t} - x_{2,t})^2}{n}}.$$

4.- CASO DE ESTUDIO: PREDICCIÓN DEL TOBT

Para el correcto análisis de los datos esenciales para el estudio y predicción de retrasos, es necesario crear e importar una batería de datos recopilados en el tiempo. En este caso, se va a proceder a crear una base de datos con los parámetros recogidos de 3 aerolíneas americanas, expresadas como 'AQ', 'AS' y 'US' con muestras obtenidas desde el 2008 en adelante.

A continuación se describen los parámetros incluidos en la base de datos:

YEAR: Año en el que se ha recopilado cada una de las muestras del estudio. En este caso todas las muestras coinciden en el año 2014, sin embargo para estudios de mayor magnitud y mayor amplitud temporal es necesario para acotar la variable temporal de los datos.

MONTH: Mes en el que se han recopilado cada una de las muestras del estudio. Este parámetro varía en una escala del 1 al 12, siendo el 1 el mes de Enero y 12 el mes de Diciembre. Como en el caso anterior todas las muestras coinciden en el mes de Enero, sin embargo para estudios de mayor magnitud y mayor amplitud temporal es necesario para acotar la variable temporal de los datos.

DAYOFMONTH: Día del mes en el que se ha recopilado cada muestra del estudio. Este parámetro varía en una escala del 1 al 31, coincidiendo con la numeración del día mensual del dato.

DAYOFWEEK: Día de la semana en el que se ha recopilado cada muestra del estudio. Este parámetro varía en una escala del 1 al 7, siendo el 1 el número equivalente al Lunes y el 7 el equivalente al Domingo. Tanto este dato como el anterior pueden ser necesarios para estimar que día de la semana (p.ej. el inicio del fin de semana) o que día del mes (p.ej. principios de mes) se estima que puede haber más congestión y por ende más retrasos en los vuelos.

DEPTIME: 'Departure time'. Hora de salida del vuelo en hora local real (LMT) expresado como un conjunto de cuatro dígitos, siendo los dos primeros la hora y los dos últimos los minutos. Tiempo computado en el momento en el que la aeronave abandona la puerta o gate. Dato esencial para calcular y prever los colapsos en función de la hora del día, horas punta y picos de trabajo.

CRSDEPTIME: 'Computerized Reservations Systems (CRS) departure time'. Tiempo que cada aerolínea tiene computado como tiempo de salida. Puede definirse como la hora de salida que la compañía estima a la que va a salir cada vuelo. Dato muy útil para comprobar la diferencia entre la hora real de salida y la hora estimada por la compañía. Indicador de la "on-time performance" de cada aerolínea.

ARRTIME: 'Arrival time'. Hora de llegada del vuelo en hora local real (LMT) expresado como un conjunto de cuatro dígitos, siendo los dos primeros la hora y los dos últimos los minutos. Tiempo computado en el momento en el que la rueda de morro de la aeronave golpea el 'Tarmac'. Dato esencial para calcular y prever los colapsos en función de la hora del día, horas punta y picos de trabajo.

CRSARRTIME: Computerized Reservations Systems (CRS) arrival time. Tiempo que cada aerolínea tiene computado como tiempo de llegada. Puede definirse como la hora de salida que la compañía estima a la que va a llegar cada vuelo. Dato muy útil para comprobar la diferencia entre la hora real de llegada y la hora estimada por la compañía. Indicador de la 'on-time performace' de cada aerolínea.

UNIQUECARRIER: Nombre en siglas de la compañía aérea de la que se recoge la muestra de datos. Dato necesario para asociar cada retraso con cada aerolínea y poder estimar y predecir que compañías producen más retrasos. Cadena de caracteres.

FLNUM: 'Flight Number'. Número de vuelo de cada muestra.

TAILNUM: 'Tail Number'. Conjunto de números y letras que componen el registro de matrícula de cada aeronave, situados en la cola del avión. Este número es único para cada aeronave y su formato depende del país de matriculación (p.ej. en el caso de EEUU está compuesto de la letra 'November' más un conjunto de dígitos). Puede encontrarse tanto en la parte posterior y anterior del avión como bajo los planos.

ACTUALELAPSEDTIME: Tiempo real de vuelo, desde la salida en origen hasta la llegada en destino, expresado en minutos como un conjunto de enteros. Dato útil para la estimación e identificación de comportamientos o retrasos en ruta.

CRSELAPSEDTIME: 'Computerized Reservations Systems (CRS) elapsed time'. Tiempo que cada aerolínea tiene computado en su sistema como tiempo de vuelo, desde la salida en origen hasta la llegada en destino, expresado en minutos como un conjunto de enteros. Dato útil para la estimación e identificación de comportamientos o retrasos en ruta y la diferencia con el tiempo real de vuelo para el cálculo de la 'on-time performance'.

AIRTIME: Tiempo real de vuelo en aire, desde la salida en origen hasta la llegada en destino, contabilizado desde que el avión abandona la pista hasta vuelve a entrar en contacto con ella, expresado en minutos como un número entero. Dato útil para la estimación e identificación de comportamientos o retrasos en ruta.

ARRDELAY: Diferencia entre la hora de llegada de CRS o estimada por la compañía y la hora real de llegada en destino, expresada en minutos como un número entero. Información útil para identificar el origen de un problema o retraso.

DEPDELAY: Diferencia entre la hora de salida de CRS o estimada por la compañía y la hora real de salida en origen, expresada en minutos como un número entero. Información útil para identificar el origen de un problema o retraso.

ORIGIN: Código del aeropuerto de origen del vuelo, expresado como una cadena de caracteres de tres letras. Código único para cada aeropuerto estandarizado por la Asociación Internacional de Transporte Aéreo (IATA).

DEST: Código del aeropuerto de destino del vuelo, expresado como una cadena de caracteres de tres letras. Código único para cada aeropuerto estandarizado por la Asociación Internacional de Transporte Aéreo (IATA).

DISTANCE: Distancia entre los aeropuertos origen y destino, expresada en millas náuticas como un número entero. Dato útil para estimar la relación entre los retrasos y la distancia de vuelo entre varios saltos.

TAXIIN: 'Taxi in time'. Tiempo de rodadura hasta puerta en el aeropuerto de destino. Expresado como un conjunto de enteros en minutos. Dato útil para identificar retrasos en las fases y operativas de rodadura.

TAXIOUT: 'Taxi out time'. Tiempo de rodadura desde la puerta hasta la pista en el aeropuerto de origen. Expresado como un conjunto de enteros en minutos. Dato útil para identificar retrasos en las fases y operativas de rodadura.

CANCELLED: Identificador de vuelo cancelado. Expresado como un valor binario, siendo cero si el vuelo se ha efectuado o uno si ha sido cancelado. Cuando este valor es uno se omiten el resto de valores temporales.

CANCELLATIONCODE: Código de cancelación. Expresado con un carácter alfabético en función de la causa de cancelación siguiendo la siguiente leyenda: A-"Carrier", B-"Weather", C-"National Air System", D-"Security".

DIVERTED: Indicador de vuelo desviado. Expresado como un valor binario, siendo cero si el vuelo se ha efectuado con normalidad o uno si ha sido desviado.

CARRIERDELAY: Retraso causado por la portadora. Expresado como un conjunto de enteros en minutos. Dato útil para identificar retrasos causados por este factor.

WEATHERDELAY: Retraso causado por inclemencias meteorológicas. Expresado como un conjunto de enteros en minutos. Dato útil para identificar retrasos causados por este factor.

NASDELAY: Retraso causado por la National Air System. Expresado como un conjunto de enteros en minutos. Dato útil para identificar retrasos causados por este factor.

SECURITYDELAY: Retraso causado por los controles de seguridad o sus inspecciones. Expresado como un conjunto de enteros en minutos. Dato útil para identificar retrasos causados por este factor.

LATEAIRCRAFTDELAY: Retraso por llegada tarde de la aeronave. Expresado como un conjunto de enteros en minutos. Dato útil para identificar retrasos causados por este factor.

4.1.- Proceso de limpieza de datos.

En este apartado se explican los pasos que hemos seguido con el fin de tratar los datos y limpiarlos para poder trabajar con estos. Es decir, se eliminarán todos aquellos datos que puedan interferir negativamente o alterar el resultado final de forma no real.

Ejemplos de esto serían, eliminar todos los valores/resultados en que el campo está vacío, ya que con esto saltarían errores. Otro ejemplo sería el de sustituir todos los valores N/A (ningún valor) por "0", para poder tratar estos datos numéricamente y omitir el error que saltaría en el momento en el que el programa encontrase el "N/A".

Para ello, primer paso que realizaremos es eliminar todos los vuelos que estén cancelados, ya que no podremos tratar los datos ni predecir retrasos en estos vuelos.

```
df = df[df.CancellationCode.isnull()]
```

Después de esto se realiza una búsqueda de todas aquellas columnas que contengan algún valor de "N/A" para, como se ha comentado anteriormente, evitar posibles alteraciones en el resultado final.

```
Se ejecuta el comando: print df.isnull().any()
```

Éste, imprimirá el listado de columnas total, indicando como "TRUE" todas aquellas columnas que contengan algún valor de "N/A". El objetivo de esto es agrupar todas estas columnas para sustituir de ellas todos los "N/A" por "0" para poder hacer un tratamiento numérico de todos los valores. Lo que obtenemos es:

```
Out[7]: Year                False
Month                False
DayofMonth           False
DayOfWeek            False
DepTime              False
CRSDepTime           False
ArrTime              True
CRSArrTime           False
UniqueCarrier        False
FlightNum            False
TailNum              False
ActualElapsedTime    True
CRSElapsedTime       False
AirTime              True
ArrDelay             True
DepDelay             False
Origin              False
Dest                 False
Distance             False
TaxiIn               True
TaxiOut              False
Cancelled            False
CancellationCode     True
Diverted             False
CarrierDelay         True
WeatherDelay         True
NASDelay             True
SecurityDelay        True
LateAircraftDelay    True
dtype: bool
```

Todas aquellas columnas marcadas de rojo, son la que contienen algún valor de "N/A", y que por tanto se deben juntar en un conjunto al que se le denominará "column_names" y con el que más tarde trataremos de apropiar todos los valores numéricamente.

Figura 32 - Impresión de variables que contienen alguna muestra 'N/A'

```
column_names =
["ArrTime", "ActualElapsedTime", "AirTime", "ArrDelay", "TaxiIn",
"CancellationCode", "CarrierDelay", "WeatherDelay", "NASDelay",
"SecurityDelay", "LateAircraftDelay"]
```

```
df[column_names] = df[column_names].fillna(0)
```

Con los comandos anteriores, lo que se ha hecho ha sido agrupar todas aquellas columnas que contenían alguna muestra con "N/A" y sustituir esta por "0".

El siguiente paso sería localizar todas las muestras numéricas que tienen un valor negativo, y se convertirá en "0", ya que incluir valores negativos en el programa de predicción también alteraría el resultado final. Para ello, muy parecido al comando explicado antes, se buscarán todas las columnas que contengan algún valor negativo, a través del comando:

```
df.where(df < 0, 0).any()
```

Los resultados obtenidos son únicamente las columnas: "ArrDelay" y "DepDelay". (Excluyendo todas aquellas columnas no numéricas).

Se eliminan todos los valores negativos de estas 2 columnas y se convierten en "0", a través de los siguientes comandos:

```
df['DepDelay'] = df['DepDelay'].where(df['DepDelay'] < 0, 0)
```

```
df['ArrDelay'] = df['ArrDelay'].where(df['ArrDelay'] < 0, 0)
```

También se elimina la columna "CancellationCode" ya que es irrelevante y no tendrá ninguna utilidad para el análisis de los datos:

```
df = df.drop("CancellationCode",1)
```

Para terminar, a modo de asegurarnos que no quede ningún dato erróneo en el fichero de datos ejecutaremos un comando con el cual hará una búsqueda de todos los "N/A" de todas las columnas y los sustituya por "0".

```
df = df.dropna()
```

Para comprobarlo se vuelve a ejecutar el comando que hace la búsqueda en todas las columnas de las muestras que contienen "N/A", e indicará como "TRUE" en caso de que si contengan algún valor, o "FALSE" en el caso de que no contenga ninguno.

```
df.isnull().any()
```



```
Out[9]: Year                False
        Month                False
        DayOfMonth           False
        DayOfWeek            False
        DepTime               False
        CRSDepTime           False
        ArrTime               False
        CRSArrTime           False
        UniqueCarrier        False
        FlightNum            False
        TailNum              False
        ActualElapsedTime    False
        CRSElapsedTime      False
        AirTime              False
        ArrDelay              False
        DepDelay              False
        Origin                False
        Dest                  False
        Distance              False
        TaxiIn                False
        TaxiOut               False
        Cancelled             False
        CancellationCode     False
        Diverted              False
        CarrierDelay          False
        WeatherDelay          False
        NASDelay              False
        SecurityDelay         False
        LateAircraftDelay    False
        dtype: bool
```

Figura 33 - Comprobación variables con 'N/A'

Como se puede comprobar, todos los resultados son "FALSE". Esto quiere decir que se ha realizado la limpieza de datos correctamente, y que estos ya están listos para ser tratados y realizar el código para la predicción del TOBT.

El último paso es guardar este nuevo fichero modificado y renombrarlo para poder trabajar con él más tarde.

```
df.to_csv("C://Users/Aitor/Desktop/TFG/BD/Cleaned_Data.csv", index=False)
```

4.2.- Pre procesamiento de datos.

Primeramente se inicia el proceso cargando el fichero, ya tratado y limpiado, que se ha guardado como 'Cleaned_Data'.

```
df = pd.read_csv("C://Users/Aitor/Desktop/TFG/BD/Cleaned_Data.csv", index_col=False)
```

Una vez cargado el fichero, listo para ser pre-procesado, se debe escoger todas aquellas variables a estudiar, que puedan influir en los tiempos del proceso. Para ello, se seleccionan todas aquellas que sean numéricas.

```
def build_features(features, data):  
  
    features.extend(['Month', 'DayOfWeek', 'DepTime', 'ArrTime',  
  
                    'Distance', 'ActualElapsedTime', 'AirTime',  
  
                    'ArrDelay', 'Distance', 'TaxiIn', 'TaxiOut',  
  
                    'CarrierDelay', 'WeatherDelay', 'NASDelay',  
  
                    'SecurityDelay', 'LateAircraftDelay'])
```

El siguiente paso sería, escoger todas aquellas variables categóricas para transformarlas a un valor numérico. Una vez hecho esto, se hará que nos devuelva la tabla con los nuevos valores a estudiar, y que imprima la cabecera de las variables a estudiar.

```
    features.append('TailNum')  
    le = LabelEncoder().fit(data['TailNum'])  
    data['TailNum'] = le.transform(data['TailNum'])  
    return data  
  
features = []  
build_features(features, df)  
print df[features].head()
```

Esto devolverá la cabecera de las primeras 5 muestras. Como se puede ver, todas aquellas muestras que tenían algún número negativo, o no válido para el estudio de las variables, aparecerán como '0', para que no influya en el resultado del estudio.

```
Out[3]:
```

	Month	DayOfWeek	DepTime	ArrTime	Distance	ActualElapsedTime	AirTime	ArrDelay	Distance	TaxiIn	TaxiOut	CarrierDelay	Weatherf
Year													
0	0	0	0.0	0.0	0	0.0	0.0	0.0	0	0.0	0.0	0.0	0.0
0	0	0	0.0	0.0	0	0.0	0.0	0.0	0	0.0	0.0	0.0	0.0
0	0	0	0.0	0.0	0	0.0	0.0	0.0	0	0.0	0.0	0.0	0.0
0	0	0	0.0	0.0	0	0.0	0.0	0.0	0	0.0	0.0	0.0	0.0
2008	1	2	825.0	1104.0	646	159.0	145.0	18.0	646	4.0	10.0	0.0	0.0

Figura 34 - Cabecera base de datos (observaciones con valores '0')

El siguiente paso para hacer más aproximada la previsión y facilitar el estudio de estos datos sería eliminar todas aquellas filas que contengan '0', ya que serían muestras no reales e influiría negativamente en el estudio a realizar. Para ello, se debe escribir el siguiente código:

```
df = df.reset_index()
df = df[df["Year"]>0]
df.head()
```

Y así se puede ver (imprimiendo la cabecera de los datos a estudiar) como desaparecen todas aquellas muestras que contenían '0'.

Out[4]:

	Year	Month	DayofMonth	DayOfWeek	DepTime	CRSDepTime	ArrTime	CRSArrTime	UniqueCarrier	FlightNum	...	Distance	TaxiIn	TaxiOut
4	2008	1	1	2	825.0	825	1104.0	1046	US	1814	...	646	4.0	10.0
5	2008	1	1	2	1925.0	1855	2155.0	2114	US	1874	...	646	10.0	16.0
9	2008	1	1	2	152.0	146	940.0	917	US	634	...	2551	4.0	14.0
12	2008	1	1	2	2206.0	2142	2345.0	2300	US	1735	...	1747	7.0	16.0
14	2008	1	1	2	1244.0	1240	1500.0	1448	US	1588	...	665	21.0	25.0

Figura 35 - Cabecera base de datos. Limpieza de datos.

Una vez realizado esto, se puede graficar todos los datos para ver las frecuencias y realizar un estudio visual de estas.

```
for x in range(0, len(features)):
    plt.figure(x)
    plt.hist(df[features].values[:,x])
    plt.title(features[x])
    plt.xlabel("Value")
    plt.ylabel("Frequency")
    plt.show()
```

Con este bucle se imprimirá un gráfico de frecuencias para cada variable, donde el vector 'X' indicará el valor y el vector 'Y' indicará la frecuencia en la que se da, cada valor.

A continuación se revisan algunos ejemplos.

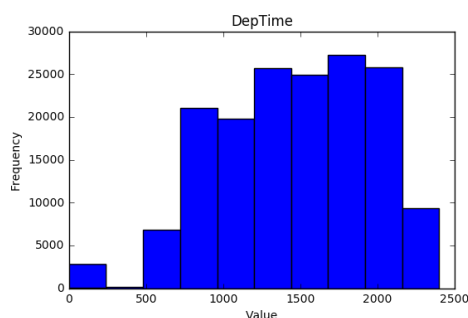
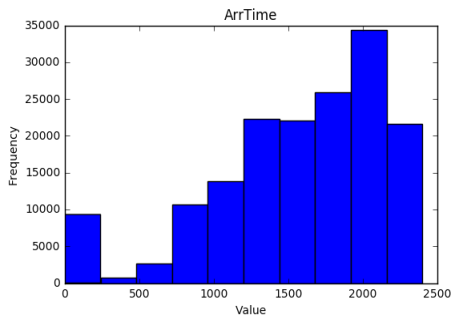


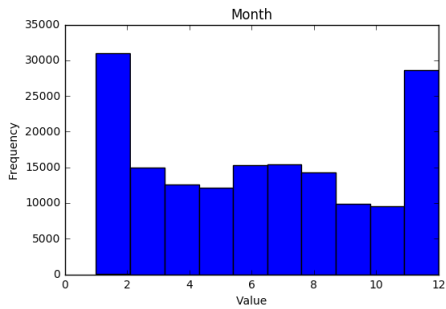
Figura 36 - Grafico de barras. Horas de salida.

En el caso de la hora de salida, la frecuencia mayor se distribuye entre las 7am y las 10pm aproximadamente, excluyendo las salidas de madrugada, ya que son horarios menos habituales para las salidas.



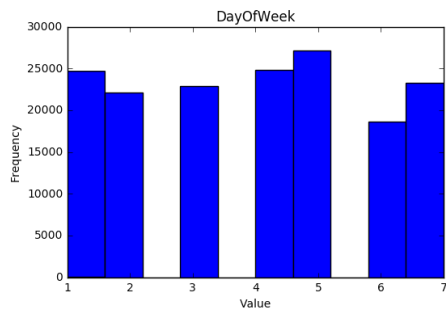
En el caso de las llegadas sin embargo, tenemos un gran número de vuelos que llegan a última hora del día. Muchos aviones de los cuales suelen repostar durante la madrugada y salen a primera hora del día siguiente.

Figura 37 - Grafico de barras. Horas de llegada.



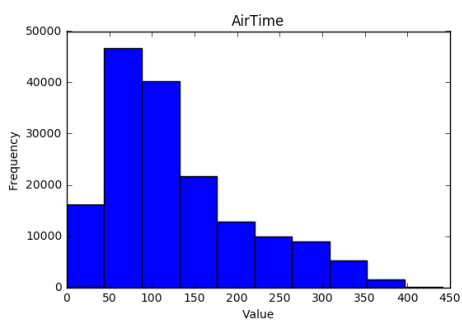
Si nos fijamos en la frecuencia de salidas, variando el mes en que se viaja, vemos como los meses que más despuntan son los meses de Enero y Diciembre (fiestas de navidad) y los meses de verano, remarcando que son los meses en que la gente viaja por vacaciones.

Figura 38 - Grafico de barras. Mes del vuelo.



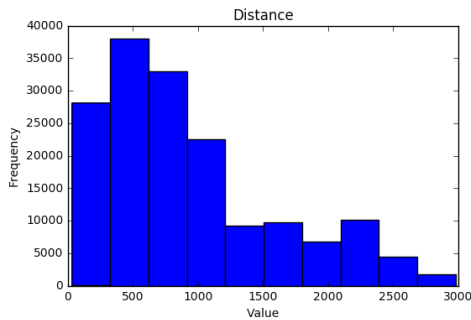
En cuanto a la frecuencia de salidas, variando el día de la semana se puede apreciar como es en viernes, con la llegada del fin de Semana el día que más se viaja, acompañado de los domingos y lunes, que se supone que serían las vueltas de estas salidas.

Figura 39 - Grafico de barras. Día de la semana.



Si graficamos la frecuencia de la variable de tiempo de vuelo, podemos ver como son más habituales los trayectos cortos que los largos. Se tiene que tener en cuenta que las aeronaves tienen un alcance de kilómetros determinados y que deben hacer escalas para repostar, en los trayectos largos.

Figura 40 - Grafico de barras. Tiempo de vuelo.



Tomando la variable de la frecuencia según la distancia, nos encontramos con el mismo caso de antes, en donde son más frecuentes las distancias cortas, y el grafico siempre está más distribuido hacia la izquierda.

Figura 41 - Grafico de barras. Distancia de vuelo.

Observando estos dos últimos gráficos nos damos cuenta que esta asimetría influirá negativamente al modelo predictivo. Es por eso que existen varios tipos de transformaciones para resolver esta asimetría o “Skewness”.

4.2.1.- Asimetría (Skewness)

Generalmente, cuando tomamos una muestra de frecuencias de cantidades, tiende a ser simétrica hacia un lado. Es decir, que hay una mayor frecuencia en un lado u otro del gráfico, ya que suele haber una mayor cantidad de valores pequeños que de valores grandes.

Si la distribución es sesgada hacia la derecha, significa que tiene un mayor número de valores grandes. En cambio, una distribución sesgada a la izquierda tiene mayor frecuencia en valores pequeños (como sería el caso de los dos últimos gráficos comentados anteriormente).

La principal razón de esta transformación de datos es eliminar la asimetría distributiva. Una distribución sesgada es una que es más o menos simétrica. Esto significa que la probabilidad de caer a cada lado de la media de la distribución es aproximadamente igual.

Una regla general a considerar es que los datos sesgados cuya relación entre el valor más alto en el menor valor es mayor que 20 tienen asimetría significativa. Además, la estadística de asimetría se puede utilizar como un diagnóstico. Si la distribución predictor es aproximadamente simétrica, los valores de asimetría estarán cerca de cero.

A medida que la distribución se desestabiliza, la estadística de asimetría se hace más grande. Del mismo modo, como la distribución se desestabiliza más a la izquierda, el valor se convierte en negativo. La fórmula para la estadística muestra asimetría es:

$$\text{skewness} = \frac{\sum(x_i - \bar{x})^3}{(n - 1)v^{3/2}}$$

$$\text{where } v = \frac{\sum(x_i - \bar{x})^2}{(n - 1)},$$

Donde x es la variable predictora, n es el número de valores, y \bar{x} es la media de la muestra del predictor.

A continuación podemos ver un ejemplo de un gráfico en su forma natural, y al lado derecho un gráfico transformado, para resolver la asimetría:

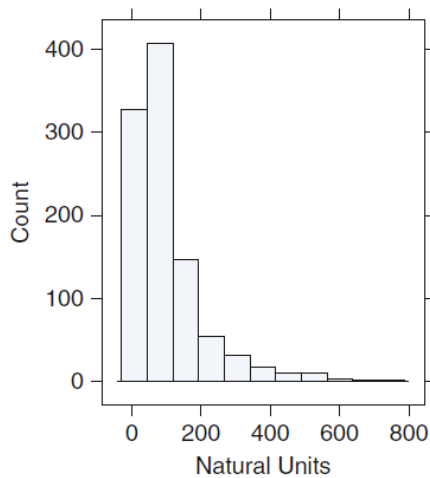


Figura 42 - 'SKEWNESS'. Ejemplo grafico sin transformar

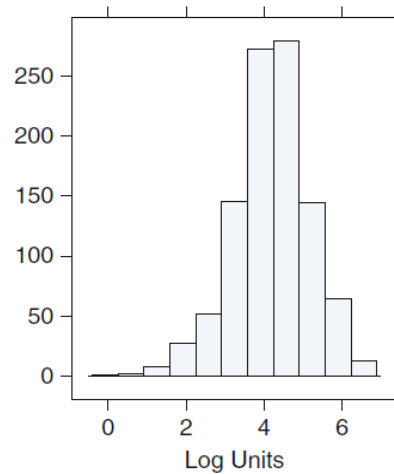


Figura 43 - 'SKEWNESS'. Ejemplo grafico transformado

En el primero se puede contemplar cómo el gráfico es asimétrico hacia la izquierda (toma mayor cantidad de valores pequeños), pero sin embargo al transformarlo, toma más la forma de una función normal, y aproximadamente el mismo número de valores altos que bajos.

Otro método para realizar un estudio generalizado de las observaciones sería el de revisar la variabilidad de estas observaciones.

Uno de los conceptos más importantes relacionados con la varianza es la desviación estándar, también conocida como típica, que representa la magnitud de la dispersión de variables de intervalo y de razón, y resulta muy útil en el campo de la estadística descriptiva. Para obtenerla, simplemente se parte de la varianza y se calcula su raíz cuadrada. La desviación estándar, por lo tanto, nos indica la variabilidad que puede tener una variable (tanto positiva como negativamente).

Es por eso que se utilizará tanto la media como la desviación estándar para medir la variabilidad de cada una de las variables, según las muestras obtenidas:

```
for x in range(0, len(features)):
    print ("%s Mean:" %features[x])
    print (df[features].values[:,x].mean(axis=0))
    print ("%s Std.Dev.:" %features[x])
    print (df[features].values[:,x].std(axis=0))
```

Con esto se despliega una lista de todas las variables indicando tanto la media, como la desviación estándar:

```

Month Mean: 6.31396855669
Month Std.Dev.: 3.53730673417
DayOfWeek Mean: 3.95934497176
DayOfWeek Std.Dev.: 1.98483609853
DepTime Mean: 1465.11000269
DepTime Std.Dev.: 492.95480073
ArrTime Mean: 1570.72097729
ArrTime Std.Dev.: 581.510490723
Distance Mean: 927.970787794
Distance Std.Dev.: 670.603404576
ActualElapsedTime Mean: 155.89461845
ActualElapsedTime Std.Dev.: 84.9639294492
AirTime Mean: 129.101677303
AirTime Std.Dev.: 81.8048383781

TaxiIn Mean: 6.37941025453
TaxiIn Std.Dev.: 4.55006436038
TaxiOut Mean: 20.5284909166
TaxiOut Std.Dev.: 14.3181489805
CarrierDelay Mean: 9.77929166972
CarrierDelay Std.Dev.: 30.078278142
WeatherDelay Mean: 0.916177657155
WeatherDelay Std.Dev.: 10.4364239032
NASDelay Mean: 8.77557519744
NASDelay Std.Dev.: 22.7562950585
SecurityDelay Mean: 0.110589500966
SecurityDelay Std.Dev.: 2.05868392614
LateAircraftDelay Mean: 12.062263136
LateAircraftDelay Std.Dev.: 30.9474695684
TailNum Mean: 256.71039023

```

Figura 44 - Output: Media + Desviación STD.

Se puede ver como generalmente, la desviación estándar de la mayoría de las variables toma unos valores muy elevados. Esto quiere decir que según todos estos datos la variabilidad es muy alta.

Teniendo en cuenta que se intenta establecer un modelo predictivo, lo más conveniente para el estudio sería que esta desviación tomase valores lo más pequeños posible. Esto significaría que la predicción sería más exacta ya que las observaciones tomadas tendrían una correlación mayor.

Lo adecuado, por lo tanto, para la realización del modelo predictivo es la transformación de estos datos para reducir esta desviación y aumentar la correlatividad entre las observaciones de cada variable.

4.2.2.- Centrado y Escalado de las variables

La transformación de datos más sencilla y común es centrar y escalar las variables predictor. Para centrar una variable predictor, el valor promedio predictor es restado de todos los valores. Como resultado de centrado, el predictor tiene una media cero. Del mismo modo, para escalar los datos, cada valor de la variable predictor es dividida por su desviación estándar. Escalar los datos coacciona los valores para tener una desviación estándar común de uno.

Por ejemplo para los elementos de una columna de datos:

$$d_{ij}(\text{autoescalado}) = \frac{(d_{ij} - \overline{d_j})}{s_j}$$

Donde d_j es el valor de la media de la columna j y s_j es la desviación estándar de los elementos de la columna j .

Se procede entonces, a la transformación de nuestro modelo a través del centrado y escalado.

```
df_scaled = preprocessing.scale(df[features])
```

Esta instrucción, hará que 'Phyton' transforme de forma automática todos los datos del 'dataframe' "features" y permita con ello un trato más efectivo de éstos.

Se puede volver a imprimir la media y la desviación estándar de todas las variables para así ver los cambios que han sufrido después de escalar y centrar los datos. Se realiza el mismo paso que se ha hecho anteriormente, simplemente indicando esta vez, que lo calcule sobre “df_scaled” que es como se le ha llamado al nuevo ‘dataframe’ que contiene los datos transformados.

```
for x in range(0, len(features)):
    print ("%s Mean:" %features[x])
    #print stats.skew(df_scaled[:,x])
    print (df_scaled[:,x].mean(axis=0))
    print ("%s Std.Dev.:" %features[x])
    print (df_scaled[:,x].std(axis=0))
```

Y los resultados obtenidos son:

Month Mean:	-1.33425467875e-16	Distance Std.Dev.:	1.0
Month Std.Dev.:	1.0	TaxiIn Mean:	-5.68100624938e-17
DayOfWeek Mean:	5.1424399077e-17	TaxiIn Std.Dev.:	1.0
DayOfWeek Std.Dev.:	1.0	TaxiOut Mean:	-1.12230276364e-16
DepTime Mean:	-1.85197329108e-16	TaxiOut Std.Dev.:	1.0
DepTime Std.Dev.:	1.0	CarrierDelay Mean:	-1.98053428878e-17
ArrTime Mean:	-1.58963936336e-16	CarrierDelay Std.Dev.:	1.0
ArrTime Std.Dev.:	1.0	WeatherDelay Mean:	-9.72894036592e-18
Distance Mean:	-5.55939449481e-17	WeatherDelay Std.Dev.:	1.0
Distance Std.Dev.:	1.0	NASDelay Mean:	2.44960819928e-17
ActualElapsedTime Mean:	-5.1424399077e-17	NASDelay Std.Dev.:	1.0
ActualElapsedTime Std.Dev.:	1.0	SecurityDelay Mean:	-3.47462155926e-18
AirTime Mean:	8.8950311917e-17	SecurityDelay Std.Dev.:	1.0
AirTime Std.Dev.:	1.0	LateAircraftDelay Mean:	1.3898486237e-18
ArrDelay Mean:	-3.61360642163e-17	LateAircraftDelay Std.Dev.:	1.0

Figura 45 - Output: Media + Desviación STD (Transformada).

Se puede ver como se han ajustado los parámetros de forma radical, y que ahora la desviación estándar es todos los casos es “1.0”. Ahora por lo tanto la variabilidad se ha reducido y por lo tanto el modelo predictivo será más eficaz.

Lo que se ha hecho con todos estos pasos ha sido acondicionar los datos, eliminando todas aquellas observaciones que no fueran determinantes, y ajustando las restantes a unos parámetros determinados para que así determinar la fuerza o el peso que tendría cada variable sobre el modelo predictivo.

4.3.- Selección de las características

Este sería el paso introductorio a la iniciación de la creación del modelo predictivo. Se trata de determinar el peso que tiene cada variable “input” sobre el resultado final “output”. Ya anteriormente hemos adecuado todos los datos para poder determinar si una variable interfiere, o no, en el retraso de la salida de la aeronave.

Esto es importante ya que dependiendo de qué variables se tengan en cuenta el resultado puede ser más eficaz o no. Se debe medir meticulosamente cuáles son aquellas variables que se deben de incluir en el modelo para que éste sea lo más aproximado posible.

Se empezará por lo tanto por definir cuál será el output, o la variable que queremos predecir como resultado, que en este caso es el retraso en la salida, que viene dado con la variable “DepDelay”.

```
delayed_minutes = df['DepDelay']
```

Una vez determinada la variable a predecir, se crea un “Random Forest” que permitirá identificar cuáles son las características que tendrán mayor impacto sobre la variable “delayed_minutes”, para así poder utilizarlas para crear el modelo predictivo.

El ‘Random forest’ es una combinación de árboles predictores tal que cada árbol depende de los valores de un vector aleatorio probado independientemente y con la misma distribución para cada uno de estos.

Para crear este “Random Forest”:

```
random.seed(111)
forest = ensemble.RandomForestRegressor(n_estimators=10,
                                         min_samples_split=2,
                                         n_jobs=-1)
forest.fit(df_scaled, delayed_minutes)
feature_importance = forest.feature_importances_
```

De donde se obtiene la variable de ‘feature_importance’, que más tarde se utilizará para medir la importancia de cada muestra, sobre el modelo predictivo.

El siguiente paso sería delimitar la máxima importancia relativa de cada variable. Por ello, se estudia el nivel de importancia mediante el porcentaje. Para ello, se maximiza la importancia de la variable “feature_importance” a ‘100’.

```
feature_importance = 100.0 * (feature_importance /
                              feature_importance.max())
```

Sobre este valor máximo, se debe indicar en el pre-procesamiento de datos sobre que umbral o porcentaje admitiremos, o no, la variable. Es necesario delimitar este umbral, ya que todas las variables tendrán una cierta importancia sobre el código de predicción y lo que se necesita es reducir a un listado en que se incluyan únicamente las variables más importantes, y así poder obtener un resultado final más preciso.

Se delimita este corte de porcentaje en 1%. Se indica por lo tanto, que únicamente se debe tener en cuenta todas aquellas variables que tengan un nivel de importancia por encima del 1%.

Se le añade un valor tan bajo, porque anteriormente se ha probado con un índice de 15% pero únicamente había una variable que lo superaba, ya que los índices de correlación no son excesivamente altos.

Para ello, se añade una nueva variable, donde se indica el umbral mínimo para que la variable sea incluida en el modelo predictivo:

```
fi_threshold = 15
```

Además, se indica que únicamente se deben seleccionar las variables que tengan el índice de importancia por encima del 'fi_threshold', dónde ya se había limitado este umbral al 1%.

```
important_idx = np.where(feature_importance > fi_threshold)[0]
```

Una vez seleccionadas las variables que cumplan con el porcentaje mínimo, hace falta que se cree una lista de estas para poder visualizar cuales son las variables que se deben incluir en el modelo.

```
features = np.array(features)
important_features = features[important_idx]
print ("\n", important_features.shape[0], "Important features(>",
        fi_threshold, "% of max importance):\n",
important_features)
```

Las variables que salen con el umbral por encima del 1% son:

```
Important features(> 1 % of max importance):
['ArrDelay' 'TaxiOut' 'CarrierDelay' 'NASDelay' 'LateAircraftDelay'
']
```

También se re-ordenará la lista de variables, según la importancia que tiene cada 'feature'.

```
sorted_idx = np.argsort(feature_importance[important_idx])
print ("\nFeatures sorted by importance (ASC):\n",
        important_features[sorted_idx])
```

Para complementar visualmente la importancia de las variables, se pueden graficar estas, mostrando el gráfico donde se muestren todas las variables junto al valor de importancia de cada una de ellas.

```
pos = np.arange(sorted_idx.shape[0]) + .5
plt.subplot(1, 2, 2)
plt.barh(pos, feature_importance[important_idx][sorted_idx],
align='center')
plt.yticks(pos, important_features[sorted_idx])
plt.xlabel('Relative Importance')
plt.title('Variable Importance')
plt.draw()
plt
```

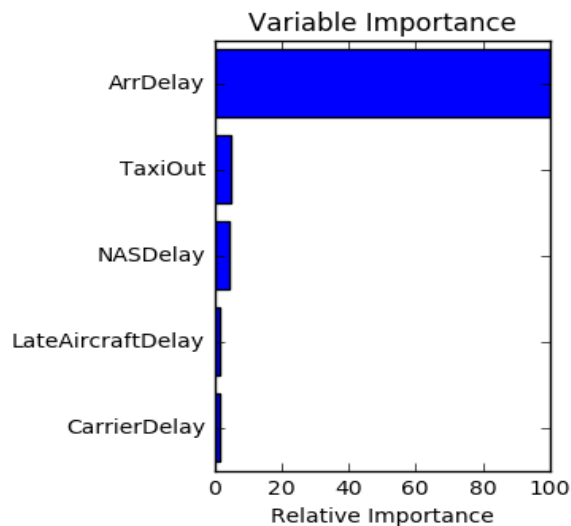


Figura 46 - Importancia relativa de las variables. Gráfico.

Con esto, en el eje 'X' se mostrará cada variable, indicándose en el eje 'Y' el índice de cada una de ellas.

Se puede apreciar, como la variable que tendría más peso sobre el modelo predictivo sería la de 'ArrDelay'. Está claro que los retrasos son una situación acumulativa en los procesos de las aerolíneas. Cuanto más tarde llegue un vuelo, más tarde saldrá y este retraso también se trasladará a la llegada del próximo vuelo del avión.

Cuando ya se ha revisado el gráfico, se imprimirán todos los índices de las variables ya que en el gráfico lo utilizamos para tener una visión general pero no se pueden apreciar los valores exactos de cada 'feature'. Para ello escribimos el siguiente código:

```
indices = np.argsort(feature_importance)[:, -1]
for f in range(df_scaled.shape[1]):
    print("(%2d) %-*s %f" % (f + 1, 30, features[f],
feature_importance[indices[f]]))
```

Y los resultados obtenidos son:

```
( 1) ArrDelay          100.000000
( 2) TaxiOut           4.752995
( 3) NASDelay          4.498018
( 4) LateAircraftDelay 1.762595
( 5) CarrierDelay      1.676409
( 6) ActualElapsedTime 0.725738
( 7) AirTime           0.676515
( 8) ArrDelay          0.630275
( 9) Distance          0.537629
(10) TaxiIn            0.516054
(11) TaxiOut           0.401263
(12) CarrierDelay      0.392366
(13) WeatherDelay      0.346891
(14) NASDelay          0.300299
(15) SecurityDelay     0.249474
(16) LateAircraftDelay 0.058269
(17) TailNum           0.021945
```

Figura 47 - Importancia relativa de las variables. Valores.

Una vez terminado esto, se pueden eliminar todas aquellas variables menos relevantes, y reordenar las variables según la importancia que tiene cada una de ellas en el modelo predictivo.

```
df_scaled = df_scaled[:, important_idx][:, sorted_idx]
df_scaled = pd.DataFrame(df_scaled)
df_scaled['DepDelay'] = delayed_minutes
important_features = np.append(important_features, ['DepDelay'])
sorted_idx = np.append(sorted_idx, len(important_features)-1)
```

Para terminar, se guardan todos los 'features' ya pre-procesados en un fichero '.csv'.

```
df_scaled.to_csv("flights_with_selected_features.csv",
header=important_features[sorted_idx], index=False)
```

Con este paso, se da fin la pre-procesamiento de datos. Se ha podido deducir cuáles serían las variables más relevantes, y que darían un resultado más óptimo incluyéndolas en el código del modelo predictivo.

4.4.- Creación del Modelo predictivo

El primer paso para empezar con la creación del modelo predictivo, es cargar en 'Phyton' el fichero, con los datos ya amoldados y reparados para la máxima eficacia posible. Para ello, se debe abrir el fichero '.csv' con el mismo nombre con el que se guardó en la parte anterior. En este caso será:

```
df = pd.read_csv("flights_with_selected_features.csv")
```

Después de esto, se debe indicar también, cuáles son las variables que, después del estudio realizado y las comparaciones correspondientes, se ha decidido que son las más influyentes para el modelo predictivo. Para ello, se crea una nueva variable donde se incluyan este grupo de 'features' influyentes, que en este caso se han seleccionado las 4 variables con más alto valor de importancia.

```
selected_features=
["ArrDelay", "TaxiOut", "NASDelay", "LateAircraftDelay"]
```

Cuando ya están las variables seleccionadas, se indicará en el modelo cuáles son los predictores y cuál es la variable a ser predicha. Es decir, se seleccionan 'inputs' y 'outputs'. En este caso, las variables a predecir ya se han nombrado antes, y tenemos una variable, creada anterioremente, que englobe a las cuatro, y para el 'output' se indica que es la variable 'DepDelay'.

```
alldata = df[selected_features]
target = df['DepDelay']
```

4.4.1.- Gradient boosting

Para empezar a crear el modelo, se seleccionan las características que deben incluirse en la predicción:

```
print ("Creating Gradient boosting model")
Gbm=
ensemble.GradientBoostingRegressor(loss='ls',n_estimators=1000,m
ax_depth=3,min_samples_split=2,verbose=True,random_state=111)
```

Después de seleccionar el modo predictivo se decide, sobre el 100% de los datos, que tanto por ciento será utilizado para el entrenamiento, y que otro tanto será utilizado para el test. Comúnmente se selecciona un 70% para el entrenamiento y un 30% para el 'testing'.

```
print ("Splitting 'alldata' into two sets: 70% Training and 30%
Testing")
train_fold,test_fold,train_y,test_y=
cross_validation.train_test_split(alldata,target, test_size=0.3,
random_state=123)
```

Una vez hecho el entrenamiento y el test, se inicia el entrenamiento para el modelo 'Gradient Boosting', donde se imprimirá una lista que mostrará el proceso de dicho entrenamiento:

```
print ("Training GB model using 'train_fold'")
m = gbm.fit(train_fold, target[train_y])
```

Training Iter	Random Forest model Train Loss	using 'train_fold' Remaining Time
1	57.1399	3.40m
2	57.1377	3.98m
3	57.1359	3.70m
4	57.1343	3.63m
5	57.1323	3.56m
6	57.1306	3.56m
7	57.1230	3.50m
8	57.1209	3.59m
9	57.1171	3.56m
10	57.1146	3.57m
20	57.0780	3.55m
30	57.0484	3.72m
40	57.0249	3.55m
50	56.9693	3.46m
60	56.9040	3.34m
70	56.8581	3.26m
80	56.8241	3.18m
90	56.7839	3.14m
100	56.7406	3.08m
200	56.4862	2.74m
300	56.3020	2.40m
400	56.1750	2.02m
500	56.0368	1.67m
600	55.8948	1.33m
700	55.7909	59.32s
800	55.6961	39.34s
900	55.6102	19.58s
1000	55.5349	0.00s

En esta tabla se muestra el proceso del 'training'. En la primera columna se puede ver el número de iteraciones que lleva procesadas. En la central, sería el recuento del Train Loss. Y en la columna derecha indica el tiempo restante hasta el final de este proceso.

Figura 48 - Proceso entrenamiento 'Gradient Boosting Model'

Una vez ejecutado el entrenamiento, se puede ejecutar el 'Testing'.

```
print ("Testing Gradient Boosting model using 'test_fold'")
predicted_y = m.predict(test_fold)
```

Una vez, ejecutada esta instrucción, se deben guardar los resultados que nos ha generado el modelo predictivo, en este caso, en un fichero '.csv' a través del cual se pueden realizar estudios de los datos o predicciones obtenidas.

```
results = pd.DataFrame()
results['Real DEP_DELAY_NEW'] = target[test_y]
results['Predicted DEP_DELAY_NEW'] = predicted_y
results.to_csv("predictionresultsGB.csv", index=False)
```

Los resultados se guardan en una base de datos, que constará únicamente de dos columnas. Una columna que mostrará el 'Real DEP_DELAY_NEW' (tiempo real de retraso) y otra columna mostrará el 'Predicted DEP_DELAY_NEW' (tiempo de retraso que se ha predicho).

Cabe recordar, que estos resultados se han obtenido de un 30% de los datos, siendo seleccionado este porcentaje aleatoriamente ya que el otro 70% se ha utilizado para el entrenamiento del modelo predictivo.

Para finalizar y medir la eficacia del modelo predictivo, se realizará un estudio de su RMSE ('Root Mean Squared Error').

```
print ("Estimating prediction error Root Mean Squared Error")
print (RMSE(target[test_y], [y for y in predicted_y]))
```

El resultado que se ha obtenido, para medir la eficacia del modelo predictivo es:

```
'Estimating prediction error Root Mean Squared Error
7.7886930346'
```

Este resultado muestra el índice de error en el modelo predictivo utilizando el sistema 'Gradient Boosting Model'. Se puede considerar un resultado bueno, pero se pueden buscar más ejemplos para buscar el valor más óptimo y decir que modelo predictivo sería el más eficaz para predecir el retraso en la salida del vuelo.

4.4.2.- 'Random Forest'

Para empezar a crear el modelo, se seleccionarán las características que se quieren incluir en la predicción:

```
print ("Creating Random Forest model")

forest = ensemble.RandomForestRegressor(n_estimators=1000,
min_samples_split=2, bootstrap=True, verbose=False,
random_state=111)
```

Después de seleccionar el modo predictivo se decide, sobre el 100% de los datos, que tanto por ciento será utilizado para el entrenamiento, y que otro tanto será utilizado para el test. Comúnmente se selecciona un 70% para el entrenamiento y un 30% para el 'testing'.

```
print ("Splitting 'alldata' into two sets: 70% Training and 30% Testing")
train_fold, test_fold, train_y, test_y =
cross_validation.train_test_split(alldata, target, test_size=0.3,
random_state=123)
```

Una vez hecho el entrenamiento y el test, se inicia el entrenamiento para el modelo 'Random Forest'.

```
print ("Training Random Forest model using 'train_fold'")
m = forest.fit(train_fold, target[train_y])
```

Ya ejecutado el entrenamiento, se puede ejecutar el 'Testing'.

```
print ("Testing Random Forest model using 'test_fold'")
predicted_y = m.predict(test_fold)
```

Después de ejecutar esta instrucción, se deben guardar los resultados que ha generado el modelo predictivo, en este caso, en un fichero '.csv' a través del cual se pueden realizar estudios de los datos o predicciones obtenidas.

```
results = pd.DataFrame()
results['Real DEP_DELAY_NEW'] = target[test_y]
results['Predicted DEP_DELAY_NEW'] = predicted_y
results.to_csv("predictionresultsRF.csv", index=False)
```

Los resultados se guardan en una base de datos, que constará únicamente de dos columnas. Una columna que mostrará el 'Real DEP_DELAY_NEW' (tiempo real de retraso) y otra columna mostrará el 'Predicted DEP_DELAY_NEW' (tiempo de retraso que se ha predicho).

Para finalizar y medir la eficacia del modelo predictivo, se estudiará su RMSE (Root Mean Squared Error).

```
print ("Estimating prediction error Root Mean Squared Error")
print (RMSE(target[test_y], [y for y in predicted_y]))
```

El resultado que se ha obtenido, para medir la eficacia del modelo predictivo es:

```
'Estimating prediction error Root Mean Squared Error
8.25464427611'
```

Pese a que el resultado del índice de RMSE podría considerarse como bueno, se puede comprobar como la comprobación realizada anteriormente con el modelo 'Gradient Boosting Model' ha dado un índice de error menor, y por lo tanto podría sacarse la conclusión que los resultados predichos con el GBM serían más aproximados que mediante la predicción con 'Random Forest'.

CONCLUSIONES

Como conclusión, tras la observación de las diferentes variables que pueden influir en el retraso del TOBT, se puede decir que el factor más importante en los tiempos de salida de un vuelo, viene a ser la hora de llegada de éste. Es obvio, que si un vuelo llega con retraso, tiene más posibilidades de salir tarde, que un avión que ha llegado a su hora, pues éste último tendrá más tiempo de margen para realizar el 'turnaround' (pese a las ayudas que se ofrecen como el 'Buffer time').

Además, se tiene que tener en cuenta que los retrasos es algo secuencial en el mundo de la aeronáutica. Un vuelo que sale con retraso, llega a destino con retraso y se van acumulando estos excesos de tiempo durante todo el transcurso de la aeronave. Es por eso, que debe irse principio de este problema para anticiparse. Un modelo predictivo sería una herramienta perfecta para adelantarse a este problema. El inconveniente que tiene un modelo predictivo de retrasos de vuelos, es la gran cantidad de variables que intervienen en los tiempos de éstos.

Es muy difícil juntar todos estos factores e implantar un modelo predictivo que pueda predecir el retraso de una llegada/salida con un 100% de exactitud. Esto se debe a que hay diversas variables muy difíciles de predecir, como podrían ser los procesos donde se puedan dar errores de factor humano, y esto supondría dificultades para la elaboración del modelo.

Es a raíz de este problema, que durante el proyecto se ha intentado estudiar el motivo principal de los retrasos, seleccionando las variables, de la base de datos de tres compañías aeronáuticas, que afectaban con más fuerza en los retrasos o tiempos de 'turnaround'. Cabe destacar la importancia de seleccionar las variables con mayor vinculación al retraso de los vuelos, ya que esto nos proporcionará mejores resultados en el modelo predictivo final.

Para ello, se ha visto como 'Phyton' es una herramienta de programación muy potente y capaz de desarrollar un modelo predictivo, previo estudio de una base de datos y selección de 'Inputs' y 'Output'. Dicho programa, ha permitido también hacer estudios más visuales y así más fáciles de entender, mediante la impresión de gráficos y diagramas.

En el estudio realizado sobre la importancia de las variables en el retraso del TOBT, se han obtenido 4 variables que tendrían mayor relevancia para la predicción de un posible retraso en la salida del vuelo. Estas variables son: 'ArrDelay', 'TaxiOut', 'NASDelay' y 'LateAircraftDelay', con unos índices de importancia de 100, 4.7, 4.4 y 1.7, respectivamente. Tanto la primera variable como la última, indican la llegada tardía desde el vuelo de origen. Está claro que cuando un avión llega con retraso, tiene más posibilidades de que salga con retraso. En cuanto a la variable 'TaxiOut', indica el tiempo de rodaje del avión hasta la pista de despegue, otro factor que también causa retrasos, debido a congestiones en la pista de despegue, que provoca tiempos de espera de las aeronaves.

En cuanto a la elaboración de modelos predictivos en este proyecto se ha conseguido elaborar, sobre una amplia base de datos constituida por información de 'turnaround' de tres aerolíneas distintas, dos modelos diferentes: 'Random Forest' y 'Gradient Boosting Model'. Ambos han dado buenos resultados. Para medir la eficacia de los modelos se ha utilizado el 'Root-Mean-Square Error' (RMSE), el cual muestra el índice de error de cada modelo. Los resultados han sido: '8,2' y '7,8' respectivamente, indicando así que el 'GBM' sería un modelo predictivo más eficaz ya que su índice de error es menor.

El proyecto ha servido también para darse cuenta de la importancia del desarrollo de las diferentes herramientas del mundo aeronáutico. Frente al auge de la utilización de estos servicios, no queda otra opción que seguir progresando para lograr un incremento en los servicios ofrecidos que vayan a la par del incremento del flujo de movimientos. Para ello, es conveniente seguir realizando estudios que puedan mejorar estos servicios y así poder satisfacer todas las necesidades de las personas dispuestas a adquirirlos.

BIBLIOGRAFÍA WEB

[1] European airline delay cost reference values

https://www.eurocontrol.int/sites/default/files/content/documents/sesar/business-case/european_airline_delay_cost_reference_values_2011.pdf

[2] CDM airports

<https://www.eurocontrol.int/articles/cdm-airports>

[3] Wikipedia - Airport Collaborative Decision Making

https://en.wikipedia.org/wiki/Airport_Collaborative_Decision_Making

[4] SlotConsulting Turnaround

<http://www.slotconsulting.eu/glossary/turnaround>

[5] Wikipedia - Aircraft ground handling

https://en.wikipedia.org/wiki/Aircraft_ground_handling

[6] The human factor and simulator training handling operators

<http://www.sciencedirect.com/science/article/pii/S092575351500020X>

[7] IATA Ground Handling Solutions

<https://www.iata.org/whatwedo/ops-infra/ground-operations/Documents/ground-handling-solutions-brochure.pdf>

[8] A-CDM in ATM and Airline Operations

<http://www.iata.org/training/acdm-atm-airline-tcvt37.aspx>

[9] Current FAA Airport Delays

<http://www.flightstats.com/go/Airport/currentAirportDelays.do>

[10] Control Delay Programs Work

<http://www.nycaviation.com/2014/06/delayed-air-traffic-control-delay-programs-work/>

[11] Ground Delays from Lightning Ramp Closures and Decision Uncertainties

<http://nldr.library.ucar.edu/repository/assets/osgc/OSGC-000-000-021-657.pdf>

[12] Wikipedia - Root-mean-square deviation

https://en.wikipedia.org/wiki/Root-mean-square_deviation

[13] Big Data Analytics automatización de modelos predictivos

<http://es.slideshare.net/dataminingperu1/big-data-analytics-automatizacin-de-modelos-predictivos>

[14] Random Forest - Berkeley Statistics - University of California

<https://www.stat.berkeley.edu/~breiman/randomforest2001.pdf>

[15] A Gentle Introduction to Gradient Boosting

http://www.ccs.neu.edu/home/vip/teach/MLcourse/4_boosting/slides/gradient_boosting.pdf

[16] Gradient Boosting Algorithm for better predictions

<https://www.analyticsvidhya.com/blog/2015/09/complete-guide-boosting-methods/>

BIBLIOGRAFÍA EDITADA

[17] Complexity Science in Air Traffic Management

Edited by Andrew Cook and Damián Rivas

[18] Airport Slots: International Experiences and Options for Reform

by Achim I. Czerny (Author)

[19] Aircraft turnaround: A guide for airport and aerodrome operators, airlines and service providers on achieving control, co-operation and co-ordination

Date of publication: 2000 ISBN: 9780717618606

ANEXO 1: DATA CLEAN.IPYNB

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import pearsonr
from sklearn.cross_validation import cross_val_score, ShuffleSplit

df = pd.read_csv("C://Users/Aitor/Desktop/TFG/BD/flightdata_AQ_AS_US.csv",index_col=0)

df = df[df.CancellationCode.isnull()]
df.isnull().any()

column_names =
["ArrTime","ActualElapsedTime","AirTime","ArrDelay","TaxiIn","CancellationCode","CarrierDelay",
"WeatherDelay","NASDelay","SecurityDelay","LateAircraftDelay"]
df[column_names] = df[column_names].fillna(0)
df.isnull().any()

df.where(df < 0, 0).any()

df[df['DepDelay'] < 0] = 0
df[df['ArrDelay'] < 0] = 0
df = df.drop("CancellationCode",1)
df = df.dropna()
df.where(df < 0, 0).any()

df.shape

df.to_csv("C://Users/Aitor/Desktop/TFG/BD/Cleaned_Data.csv",index=False)
```

ANEXO 2: PREPROCESING.IPYNB

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
from scipy.stats import pearsonr
from sklearn.cross_validation import cross_val_score, ShuffleSplit
from sklearn import preprocessing
from sklearn.preprocessing import LabelEncoder
from sklearn import preprocessing
from sklearn import ensemble
import scipy.stats as stats
import random

df = pd.read_csv("C://Users/Aitor/Desktop/TFG/BD/Cleaned_Data.csv",index_col=0)

def build_features(features, data):
    features.extend(['Month','DayOfWeek','DepTime','ArrTime',
                    'Distance', 'ActualElapsedTime', 'AirTime',
                    'ArrDelay', 'Distance', 'TaxiIn', 'TaxiOut',
                    'CarrierDelay', 'WeatherDelay', 'NASDelay',
                    'SecurityDelay', 'LateAircraftDelay'])
    features.append('TailNum')
    le = LabelEncoder().fit(data['TailNum'])
    data['TailNum']=le.transform(data['TailNum'])

    return data

features = []
build_features(features, df)
df[features].head()

df = df.reset_index()
df = df[df["Year"]>0]
df.head()

for x in range(0, len(features)):
    plt.figure(x)
    plt.hist(df[features].values[:,x])
    plt.title(features[x])
    plt.xlabel("Value")
    plt.ylabel("Frequency")
    plt.show()

for x in range(0, len(features)):
    print ("%s Mean:" %features[x])
    print (df[features].values[:,x].mean(axis=0))
    print ("%s Std.Dev.:" %features[x])
    print (df[features].values[:,x].std(axis=0))

df_scaled = preprocessing.scale(df[features])
```

```

for x in range(0, len(features)):
    print ("%s Mean:" %features[x])
    #print stats.skew(df_scaled[:,x])
    print (df_scaled[:,x].mean(axis=0))
    print ("%s Std.Dev.:" %features[x])
    print (df_scaled[:,x].std(axis=0))

delayed_minutes = df['DepDelay']

random.seed(111)
forest = ensemble.RandomForestRegressor(n_estimators=100,
                                         min_samples_split=2,
                                         n_jobs=-1)
forest.fit(df_scaled, delayed_minutes)
feature_importance = forest.feature_importances_

feature_importance = 100.0 * (feature_importance /
                              feature_importance.max())

fi_threshold = 1

important_idx = np.where(feature_importance > fi_threshold)[0]

features = np.array(features)
important_features = features[important_idx]
print ("\n", important_features.shape[0], "Important features(>",
        fi_threshold, "% of max importance):\n", important_features)

sorted_idx = np.argsort(feature_importance[important_idx])
print ("\nFeatures sorted by importance (ASC):\n",
        important_features[sorted_idx])

%matplotlib inline
pos = np.arange(sorted_idx.shape[0]) + .5
plt.subplot(1, 2, 2)
plt.barh(pos, feature_importance[important_idx][sorted_idx], align='center')
plt.yticks(pos, important_features[sorted_idx])
plt.xlabel('Relative Importance')
plt.title('Variable Importance')
plt.draw()
plt

indices = np.argsort(feature_importance)[::-1]
for f in range(df_scaled.shape[1]):
    print("(%2d) %-*s %f" % (f + 1, 30, features[f], feature_importance[indices[f]]))

df_scaled = df_scaled[:, important_idx][:, sorted_idx]

```

```
df_scaled = pd.DataFrame(df_scaled)
df_scaled['DepDelay'] = delayed_minutes
important_features = np.append(important_features,['DepDelay'])
sorted_idx = np.append(sorted_idx,len(important_features)-1)

df_scaled.to_csv("flights_with_selected_features.csv",
                 header=important_features[sorted_idx],index=False)
```

ANEXO 3: PREDICTIVE MODEL.IPYNB

```
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
import statsmodels.api as sm
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
from sklearn import ensemble
from sklearn import cross_validation

def RMSE(y, yhat):
    rmspe = np.sqrt(np.mean((y - yhat)**2))
    return rmspe
df = pd.read_csv("flights_with_selected_features.csv")
df = df.replace(np.nan,0,regex=True)

selected_features = ["ArrDelay","TaxiOut","NASDelay","LateAircraftDelay"]
alldata = df[selected_features]
target = df['DepDelay']

print ("Creating Random Forest model")
forest = ensemble.RandomForestRegressor(n_estimators=1000, min_samples_split=2,
bootstrap=True, verbose=False, random_state=111)
#gbm =
ensemble.GradientBoostingRegressor(loss='ls',n_estimators=1000,max_depth=3,min_samples
_split=2,verbose=True,random_state=111)

print ("Splitting 'alldata' into two sets: 70% Training and 30% Testing")
train_fold, test_fold, train_y, test_y = cross_validation.train_test_split(alldata, target,
test_size=0.3, random_state=123)

print ("Training Random Forest model using 'train_fold'")
m = forest.fit(train_fold, target[train_y])
#m = gbm.fit(train_fold, target[train_y])

print ("Testing Random Forest model using 'test_fold'")
predicted_y = m.predict(test_fold)

results = pd.DataFrame()
results['Real DEP_DELAY_NEW'] = target[test_y]
results['Predicted DEP_DELAY_NEW'] = predicted_y
results.to_csv("predictionresults.csv", index=False)

print ("Estimating prediction error Root Mean Squared Error")
print (RMSE(target[test_y], [y for y in predicted_y]))

plt.plot(target[test_y].iloc[0:100], marker='o', linestyle='--', color='b')
plt.plot(predicted_y[0:100], marker='o', linestyle='--', color='r')
plt.legend(['Real Departure Delays', 'Predicted Departure Delays'], loc='upper left')
plt.show()
```