

Predicción de la velocidad del viento mediante un modelo *GradientBoostingTree* de *machine learning* aplicable en la gestión del tránsito aéreo

PREDICTION OF WIND SPEED BY A GRADIENT BOOSTING TREE MODEL OF MACHINE LEARNING APPLICABLE TO AIR TRAFFIC MANAGEMENT

Alexander Reyes González, areyesg@libertadores.edu.co¹

José John Fredy González Veloza, jjgonzalezv02@libertadores.edu.co²

Facultad de Ingeniería y Ciencias Básicas, Fundación Universitaria Los Libertadores

RESUMEN

Mediante el presente estudio se identificó al *Gradient Boosting Tree* (árbol de potenciación de gradiente) como el mejor modelo estadístico dentro de un grupo de diez, cuyo diseño estuvo orientado a la predicción de la intensidad del viento a diferentes alturas con aplicación en la gestión del tránsito aéreo y particularmente en las actividades de diseño de rutas y procedimientos de vuelo propias de la organización y gestión del espacio aéreo. Se analizaron datos de radiosonda registrados entre los años 2015 y 2020, obteniendo un valor de error cuadrático medio de 13,6 nudos para la totalidad de datos correspondientes a las alturas entre 2.546 m (8.353 ft) y 36.978 m (106.442 ft). Al aplicar el modelo a los datos con límite de altura de 15.240 m, el error cuadrático medio se redujo a 8,78 nudos; no obstante, los resultados de las predicciones mostraron una tendencia común hacia un valor aproximado de 15 nudos, con lo cual se determinó que no son útiles en la práctica frente a la existencia de valores de viento superiores. La solución a este inconveniente se encontró en la generación de una nueva base de datos a partir de registros únicos de altura y la asignación de valores críticos (máximos y mínimos) extraídos de cada variable predictora. Una vez sometidos los datos al modelo, se obtuvo un RMSE de 9,24 nudos. Finalmente, el modelo permitió la predicción de valores de viento para altitudes entre 8.400 ft y 30.000 ft, con un intervalo de 100 ft y un total de 217 registros. Se observó que el modelo resulta de gran utilidad en la determinación de la probabilidad de ocurrencia de los datos predichos, facilitando de esta manera la determinación de un parámetro técnico exigido para considerar que el valor de velocidad del viento es utilizable con fines de diseño de procedimientos de vuelo; esto es, una probabilidad de 2σ (95 %) correspondiendo a un valor de 43,05 nudos.

Palabras clave: Radiosonda, viento, *machine learning*, probabilidad, gestión del tránsito aéreo.

ABSTRACT

Through this study, the Gradient Boosting Tree was identified as the best statistical model within a group of ten, which design was aimed at predicting wind intensity at different heights, for application in air traffic management, and particularly in the design of routes and flight procedures as activities related to airspace organization and management. Rawinsonde data recorded between 2015 and 2020 were analyzed, obtaining a mean square error value of 13.6 knots for all data corresponding to heights between 2,546 m (8,353 ft) and 36,978 m (106,442 ft). When applying the model to the data with a height limit of 15,240 m, the root mean square error was reduced to 8.78 knots; however, the results of the predictions showed a common tendency towards an approximate value of 15 knots, with which it was determined that they are not useful in practice in the presence of higher wind values. The solution to this problem was found in the generation of a new database from unique height records and the assignment of critical values (maximum and minimum) extracted from each predictor variable. Once the data were

processed by the model, an RMSE of 9.24 knots was obtained. Finally, the model allowed the prediction of wind values for altitudes between 8,400 ft and 30,000 ft, with an interval of 100 ft and a total of 217 records. It was observed that the model is very useful in determining the probability of occurrence of the predicted data, thus facilitating the determination of a technical parameter required to consider that the value of wind speed is usable for flight procedures design purposes; that is, a probability of 2σ (95%) corresponding to a value of 43.05 knots.

Keywords: Rawinsonde, wind, machine learning, likelihood, air traffic management.

1. INTRODUCCIÓN

La Organización de Aviación Civil Internacional (OACI), sustenta que las expectativas de la sociedad y la comunidad internacional de aviación relacionadas con la capacidad para satisfacer la demanda presente en condiciones de máximo tránsito, la eficiencia operacional y rentabilidad económica asociada con la posibilidad de seguir trayectorias de vuelo preferentes y la flexibilidad para los usuarios del espacio aéreo de modificar las trayectorias de vuelo de manera dinámica, serán satisfechas en la medida en que los Estados logren, mediante estrategias de respuesta pertinentes y oportunas, el nivel de actuación requerido en sus sistemas de navegación y de gestión del tránsito aéreo (*Air Traffic Management [ATM]*) (OACI, 2008, pp.E-1 – E-2).

La ATM se basa en la participación de siete componentes dentro de los cuales se encuentra la organización y gestión del espacio aéreo (*Airspace Organization and Management [AOM]*) para el establecimiento de estructuras del espacio aéreo que dan cabida a los distintos tipos de actividades aéreas, volúmenes de tránsito y diversos niveles de servicios; todo enmarcado en los principios de economía, eficiencia y servicios integrados (OACI, 2005, p.2-1). Esta integración y disponibilidad de servicios ATM como producto de la participación conjunta de los componentes del sistema de gestión de tránsito aéreo, puede consolidarse únicamente en un entorno donde la disponibilidad e intercambio de información se lleva a cabo sin limitaciones. Esto facilita el análisis de datos para la toma de decisiones colaborativa y el modelado de herramientas que soportan las iniciativas de los Estados para hacer frente a los retos derivados de la dinámica particular del transporte aéreo.

La AOM puede ver materializado un reto en la provisión del servicio de diseño de procedimientos de vuelo relacionado con el trazado de trayectorias para las aeronaves en las fases de salida, ruta, espera¹, llegada y aproximación; la optimización del espacio aéreo para dar cabida a un mayor número de aeronaves y la organización del espacio aéreo para facilitar la aplicación de procedimientos asociados con el servicio de control de tránsito aéreo; todo esto con la garantía de seguridad operacional ante la presencia de obstáculos físicos (naturales y artificiales), otras aeronaves y limitaciones del espacio aéreo, como por ejemplo aquellas áreas restringidas o de operación militar.

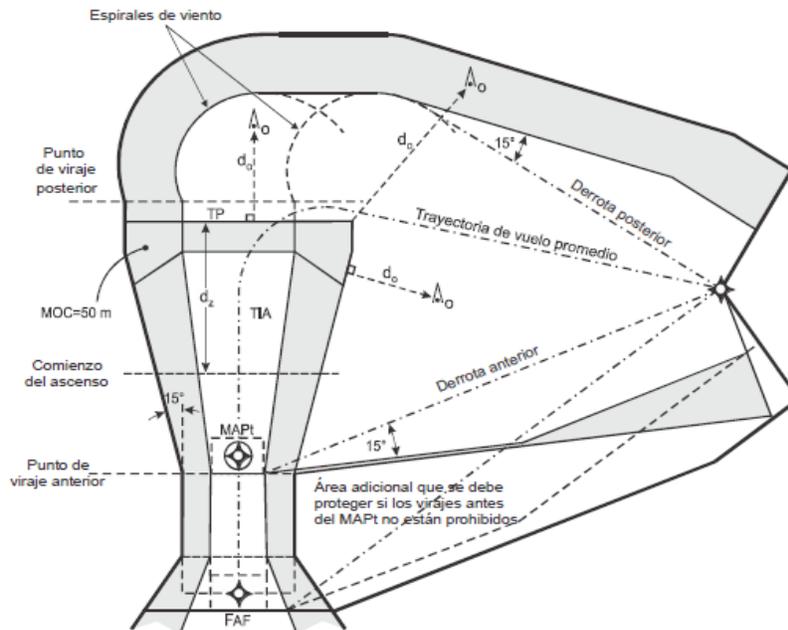
Las trayectorias de vuelo de las aeronaves a lo largo de las rutas y procedimientos se encuentran asociadas a un área de evaluación de obstáculos cuyo trazado puede requerir datos de viento, particularmente para aquellas partes donde la aeronave realiza un viraje; por ejemplo, en la aproximación (Imagen 1) donde se utilizan espirales de viento y los procedimientos de espera (Imagen 2), casos en los cuales el área de protección resultante guarda relación con el radio de viraje calculado en función del viento y la velocidad de las aeronaves, el cual es un parámetro que determina la trayectorias de vuelo posibles² hacia los puntos de referencia para la navegación.

¹ El procedimiento de espera se relaciona con el vuelo de una aeronave describiendo una geometría de hipódromo con referencia a un punto geográfico, en el cual la tripulación espera una autorización del control de tránsito aéreo para continuar su vuelo, o por razones técnicas de la aeronave, meteorológicas o de disponibilidad de la infraestructura en tierra.

² Identificados en la Imagen 1 como Derrota posterior, Derrota anterior y Trayectoria de vuelo promedio.

Imagen 1

Área de protección de un procedimiento de aproximación frustrada



Fuente: (OACI, 2020, p.III-2-2-17)

El hallazgo de obstáculos dentro de los límites de las áreas³ significa la asignación de un nivel mínimo de vuelo para garantizar la separación vertical y eventualmente, los límites laterales de las áreas de protección motivan la necesidad de modificar la trayectoria de vuelo para alejar su límite exterior de las zonas restringidas, un límite fronterizo o un obstáculo o grupo de obstáculos.

Frente a la ausencia de valores estadísticos de la intensidad del viento, los Estados deben adoptar un valor estandarizado por la OACI mediante la siguiente ecuación:

$$\text{Viento OACI} = w = 2h + 47 \quad (1)$$

Donde w es el viento, h es la altitud de la aeronave en miles de pies y 47 es una constante cuya unidad es el nudo. De esta manera, para una aeronave que realice un viraje entre dos tramos de vuelo o en un procedimiento de espera a una altitud de 17.000 pies, es necesario diseñar un área de protección basada en un viento de 81 nudos, valor que, sumado a la velocidad de las aeronaves, determinan las dimensiones del área de evaluación de obstáculos, las cuales resultan exageradas frente a un valor de viento estadístico.

Si para ese nivel se obtiene un valor estadístico de 32 nudos, la diferencia entre las áreas de protección para diferentes procedimientos de espera al mismo nivel se observa en la Imagen 2.

Las consecuencias de evaluar obstáculos con un área de mayores dimensiones son:

- Identificación de un número mayor de obstáculos y la eventual necesidad de asignar niveles de vuelo más altos, perdiendo de esta manera la disponibilidad de niveles que pueden ser utilizados por otras aeronaves.
- Solapamiento entre áreas de diferentes procedimientos, lo que obliga a la asignación de diferentes niveles de vuelo a las aeronaves involucradas, perdiendo la posibilidad de mantenerlas a un mismo nivel.

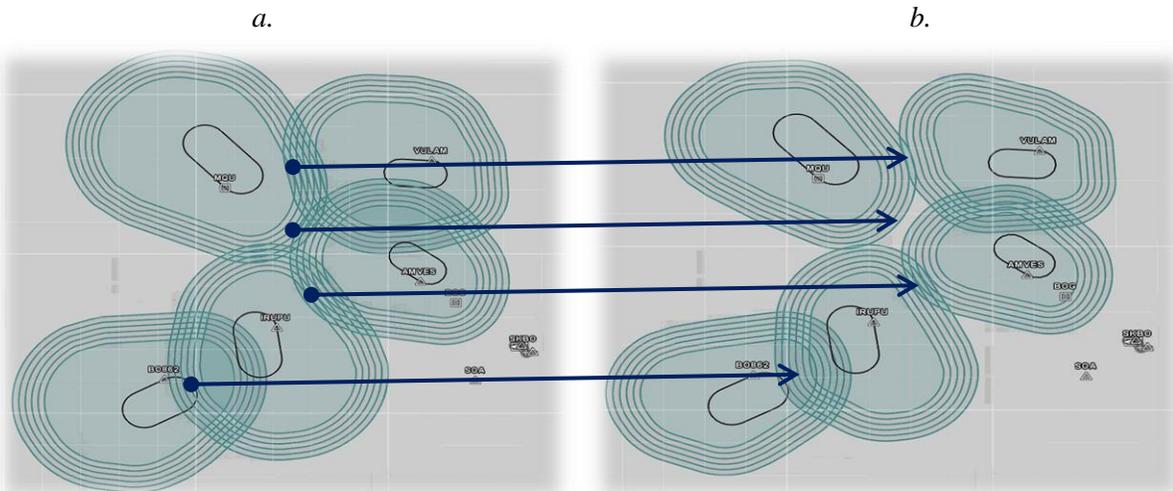
³ En la Imagen 1 los obstáculos se identifican mediante la letra "o" y flechas con líneas a trazos a una distancia d_o .

- Solapamiento entre áreas de las rutas, procedimientos y los límites laterales de espacios aéreos restringidos y límites fronterizos, escenarios que demandan la modificación de las trayectorias de vuelo o la asignación de niveles de vuelo no eficientes.
- Identificación de obstáculos que motivan la asignación de niveles restrictivos en función de la probabilidad de lograr un aterrizaje en condiciones meteorológicas o de obstáculos adversas.

Imagen 2

Reducción de las áreas de protección de los patrones de espera:

a. Con valor de viento normalizado de 81 nudos, b. Con valor de viento estadístico de 32 nudos



Nota: Elaboración propia con software FPDAM (*Flight Procedure Design and Airspace Management*).

Con la intención de mitigar los efectos de las áreas de protección con dimensiones exageradas, particularmente en el caso de los procedimientos de espera, algunos Estados han adoptado plantillas de criterios de diseño promulgados por otros Estados. Esto, al identificar una reducción en las dimensiones de las áreas utilizadas para la evaluación de obstáculos; no obstante, se desconoce la pertinencia de esta medida, teniendo en cuenta que no existe una herramienta que permita determinar si las plantillas adoptadas dan cuenta del comportamiento del viento originado por las condiciones locales, convirtiéndose en un tema de interés desde el punto de vista de la seguridad operacional. Estas situaciones impiden que los Estados alcancen el nivel de actuación ATM requerido, siendo consecuente la dificultad para satisfacer las expectativas de los usuarios con relación a la capacidad del espacio aéreo, la eficiencia operacional mediante trayectorias de vuelo rentables y preferentes, y la probabilidad de acceso a los aeródromos que presentan condiciones de terreno y meteorológicas complejas.

2. JUSTIFICACIÓN

Autores como Cole et al. (1999) informan sobre estudios relacionados con la precisión de datos meteorológicos (viento y temperatura) registrados por aeronaves y radiosondas, los cuales tienen injerencia en la ATM, particularmente como soporte a la medición de la productividad de las áreas de control terminal⁴ (*Terminal Control Area [TMA]*) (pp.1022-1023). De manera análoga, Kim et al. (2015), describen los beneficios para la ATM derivados de la metodología utilizada para identificar las rutas óptimas con base en estudios del viento

⁴ Una área de control terminal – TMA es un espacio aéreo donde se suministra el servicio de control de tránsito aéreo a las aeronaves que evolucionan desde y hacia uno o más aeródromos.

(*Wind-optimal Route [WOR]*); en la cual, a partir de datos de viento obtenidos de los modelos de pronóstico meteorológico, se logró el establecimiento de rutas laterales para evitar áreas de turbulencia entre los aeropuertos de Los Ángeles y Nueva York (p.766).

Con base en los anteriores antecedentes, se identifica que el modelo estadístico propuesto para predecir la intensidad del viento a diferentes alturas, representa una solución oportuna para los diseñadores de procedimientos de vuelo quienes idealmente deben tener acceso a estos valores para cualquier nivel de vuelo, con el fin de trazar áreas de protección ajustadas a las características locales del viento; además de facilitar su vez la determinación de viabilidad para la implementación de criterios de diseño promulgados por otros Estados. Adicionalmente, esta herramienta estadística se consolida como una base para ampliar los estudios de viento a partir de datos georreferenciados registrados por las aeronaves en vuelo, con lo cual será posible determinar la resolución espacial en el plano horizontal y perfil vertical de las predicciones a lo largo de rutas y procedimientos que resulten de interés particular para los fines de la ATM.

En conclusión, el uso de un modelo estadístico para predicción de valores de intensidad del viento a diferentes alturas a partir del registro de datos observados permite optimizar el uso del espacio aéreo, aumentar la capacidad de los aeródromos y garantizar la reducción en el consumo de combustible. Esto se traduce en una contribución significativa para el aumento del nivel de actuación de la ATM respecto a la satisfacción de las expectativas de la OACI y los usuarios del espacio aéreo con relación a los principios de seguridad operacional, capacidad, eficiencia y sostenibilidad ambiental.

3. MARCO TEÓRICO

3.1 El viento

3.1.1 Definición

La atmósfera terrestre manifiesta un constante intento por reducir las diferencias causadas por la distribución irregular de energía en la que participan la humedad y la presión. Los consecuentes procesos físicos de autorregulación involucran a su vez a la temperatura, cuya diferencia entre los polos y el ecuador facilita la transformación de energía que origina la circulación atmosférica del planeta. Este paso de energía calorífica a cinética genera el movimiento de las masas de aire sobre las cuales actúan otras fuerzas como la gravedad, la presión el rozamiento y la fuerza de Coriolis; cuya sumatoria dan origen al viento cuando predomina el movimiento horizontal (Alarcón y Castillo, 1999, pp. 63 – 85).

Con relación a sus características generales, Crespo y Zúñiga (2015) indican:

El viento es una magnitud vectorial y para caracterizarlo hay que medir su dirección y su módulo. El viento de superficie se mide a la altitud normalizada de 10 metros con la ayuda de un anemómetro para la velocidad y una veleta para la dirección. La velocidad del viento, que también se llama intensidad o “fuerza”, se mide en el SI en m/s, pero es frecuente medirlo en nudos. Un nudo es una milla marina por hora que equivale a $1850 \text{ m/h} \approx 0,5 \text{ m/s}$ (p. 108).

3.1.2 Características del viento en Colombia

De acuerdo con la descripción hecha por Ruíz et al. (2017), la ubicación geográfica de Colombia relativa al Ecuador terrestre hace que el viento local sea influenciado por los vientos alisios del noreste en el hemisferio norte y del sureste en el hemisferio sur. Adicionalmente, debido a la débil o casi nula fuerza de Coriolis, su comportamiento está determinado en gran medida por la cordillera de los Andes, la predominancia de los sistemas de baja presión y la presencia de los océanos Atlántico y Pacífico.

El viento local también presenta variaciones a causa del desplazamiento de la Zona de Confluencia Intertropical (ZCIT) a lo largo del año, presentando baja intensidad en el área de encuentro de los alisios y alta a mayores distancias de esa zona (pp.30 – 31). La topografía colombiana presenta áreas protegidas del viento y áreas caracterizadas por fuertes ráfagas. Las montañas de gran altura presentan un flujo de viento fuerte, caliente y seco, que se desarrolla en las laderas de sotavento⁵ de la cordillera central, así como en las regiones Caribe, Andina y piedemonte amazónico, inmediaciones de Salazar, Cúcuta, San Isidro, Sierra Nevada de Santa Marta, altiplano cundiboyacense, sectores localizados entre el alto Magdalena y el piedemonte amazónico, y en la Serranía de San Lucas. La variación diurna del viento en superficie también juega un rol muy destacado en las circulaciones locales como la brisa de mar-tierra y la brisa valle-montaña, las cuales se desarrollan como resultado del calentamiento desigual de la tierra y el mar o de las laderas y los valles en zonas montañosas (Ruíz et al., pp.31 – 32).

3.2 Adquisición de datos de viento en altura

3.2.1 Medios convencionales

a. Radiosondas

Una radiosonda meteorológica es una herramienta electrónica de medición que proporcionará datos del perfil vertical para variables atmosféricas como temperatura, humedad relativa, presión atmosférica, dirección y velocidad del viento. La componen una serie de sensores electrónicos y un módulo transmisor de radio Frecuencia, cuyo funcionamiento es integrado y regulado mediante microcontroladores. La radiosonda es enviada a la atmósfera generalmente mediante un globo de helio, y una vez el microcontrolador comienza a recibir las variables, las junta en una señal y las envía a un receptor localizado en tierra. Los datos obtenidos son utilizados para diversos fines: estudios espaciales, diseño y construcción de edificios de gran altura, la contaminación del aire, el pronóstico del tiempo y el comportamiento general de la atmósfera (Sánchez, 2016).

Durante el año 2015 se realizó un estudio de la atmósfera Ecuatorial con base en datos de presión, temperatura, humedad relativa, punto de rocío velocidad y dirección del viento obtenidos mediante el lanzamiento de 5 radiosondas, el cual permitió identificar el perfil vertical del ozono, de las variables físicas atmosféricas, el valor de altura de la tropopausa y la tasa de enfriamiento adiabática de la atmósfera. La contribución principal del estudio se resume en parámetros de referencia para la validación de modelos atmosféricos y la optimización de las predicciones de las variables meteorológicas (Salvador, 2015, pp.5-16). Una de las principales conclusiones del estudio indica que el perfil vertical de la velocidad del viento aumenta con la altitud, debido a la reducción de la fricción a mayores altitudes. El aumento no es lineal debido a la fricción que existe entre capas de la atmósfera y por el fenómeno de la mezcla de las masas de aire. La tropósfera de Quito se caracteriza por fuertes corrientes de aire, alcanzando 97 km/h a 15 km de altitud (p. 50).

3.2.2 Medios no convencionales

a. Aeronaves comerciales

Cole et al. (1999) documentan uno de los primeros antecedentes sobre el uso de aeronaves de transporte aéreo comercial para la adquisición de datos de viento en las áreas occidental y central de Estados Unidos. Esto se logró mediante la configuración del sistema de direccionamiento y reporte de comunicaciones de la aeronave (*Aircraft*

⁵ Barlovento es el punto en el horizonte desde donde el viento sopla. Sotavento es el punto opuesto.

Communications, Addressing, and Reporting System [ACARS]) para estimar la desviación estándar del error asociado con los registros de viento y temperatura. Los datos son obtenidos mediante comparación entre el vector de movimiento de la aeronave respecto a la tierra suministrado por los sistemas inerciales de navegación (*Inertial Navigation Systems [INS]*)⁶ y el vector de movimiento de la aeronave respecto a la masa de aire suministrado por sensores de velocidad del aire y del rumbo. El error horizontal del viento se observó en $1,1 \text{ m/s}$ atribuido a las maniobras de la aeronave; no obstante, se considera que estas observaciones resultan más precisas que las obtenidas a través de radiosondas meteorológicas, lo cual ha permitido el uso de estos datos directamente en los sistemas de pronóstico operacional (p.1032 – 1033).

Un testimonio adicional sobre el uso de aeronaves para la adquisición de datos meteorológicos lo suministran Ikuchi et al. (2018) al relacionar los datos de viento con la precisión de los modelos numéricos de predicción del clima y su impacto positivo en la ATM, dado su alcance en términos de seguridad operacional y eficiencia permitiendo la selección de la mejor ruta de vuelo bajo el concepto de operaciones de vuelo basadas en la trayectoria (*Trajectory-based Operations [TBO]*) (pp.335-336).

b. Drones

Como alternativa a las radiosondas tradicionales, en Quito – Ecuador se llevó a cabo la implementación exitosa de un sistema de tratamiento de datos meteorológicos soportado en la operación de un dron, cuyo estudio estuvo orientado a la generación artificial de nubes. El trabajo realizado permitió la adquisición y almacenamiento de información meteorológica como insumo para estudios atmosféricos en áreas geográficas pequeñas cercanas al volcán Pululahua y San Antonio de Pichincha. El sistema se caracterizó por contar con un módulo de monitoreo y control permanente, un componente aéreo (móvil) y otro fijo en tierra. La parte móvil para adquisición de información meteorológica en vuelo mediante sensores de parámetros meteorológicos, dispositivos GPS, tarjetas de comunicación inalámbrica y un Raspberry Pi, para su posterior transmisión hacia el sistema en tierra (Chango et al., 2017). Las pruebas de funcionamiento se realizaron llevando al sistema móvil a diferentes alturas con el fin de apreciar el comportamiento de la presión, humedad y temperatura en función de la altitud.

c. Teléfonos móviles

Según lo indica Waller (2020), gracias a los avances en las técnicas de computación, los modelos de pronóstico meteorológico basados en métodos numéricos pueden mejorar su resolución temporal y espacial mediante la incorporación de datos suministrados por el público en general a través de estaciones meteorológicas automáticas o una aplicación instalada en los teléfonos móviles. Advierte que a pesar de su potencial aporte, es imperativa la evaluación de la calidad para determinar el nivel de utilidad de los datos adquiridos mediante estas tecnologías. Esta oportunidad fue visualizada durante el “*Big data assimilation workshop*”, evento que fue liderado en el año 2019 por la *Royal Meteorological Society* (p.1).

3.3 Procesamiento y utilización de los datos de viento

3.3.1 Modelos aeronáuticos de pronóstico

De acuerdo con la OACI (2019), un pronóstico se define como la declaración de las condiciones meteorológicas previstas en un aeródromo, sobre cierta área o a lo largo de una ruta. Se ve limitado por la variación de las

⁶ Un sistema de navegación inercial calcula la posición, velocidad y dirección de una aeronave independientemente de fuentes externas de navegación, a partir de sensores de rotación y aceleración instalados a bordo.

magnitudes espacio/tiempo y por la naturaleza de algunos fenómenos meteorológicos como el viento en la superficie. El valor pronosticado de cualquiera de los elementos se entiende como el valor más probable durante el período indicado como vigencia y se espera una precisión de 10 nudos para el 90% de la veces (p. 3-1).

Las herramientas de pronóstico meteorológico existentes son modelos fisicomatemáticos que procesan datos de la atmósfera para proyectar el comportamiento de las variables dentro de un lapso determinado. Este tipo de modelos se valen de la aplicación de métodos numéricos que involucran cálculos complejos relacionados con la mecánica de fluidos, termodinámica y conceptos derivados de la estocástica para manipular grandes volúmenes de datos en poco tiempo y con validez espacial con alcance global o regional.

3.3.2 Modelos de predicción

a. Redes neuronales

Cattani et al. (2004), evaluaron la calidad de la clasificación realizada por redes neuronales con relación a la predicción de la velocidad del viento y ráfagas de viento, con una validez entre 1 y 24 horas. Las predicciones se realizaron con base en series de tiempo e información de modelado locales. La selección de las características de entrada fue iniciada por las series de tiempo y completada mediante comparación empírica y clasificación por redes neuronales de avance que fueron entrenadas con varios conjuntos de entradas. Las variables incluidas en los conjuntos de entrada fueron la hora del día, día del año, observación individual del viento y otras características extraídas del modelo de datos. Las redes neuronales de clasificación demostraron una calidad superior comparadas con resultados de pruebas de persistencia con relación a una medida de rendimiento específica (p.727). No obstante, los estudios citados como referencias no indican la manera de gestionar adecuadamente la falta de balance en los datos y la ambigüedad resultante de observaciones con características altamente aleatorias.

b. Métodos múltiples

Uno de los estudios recientes relacionados con la predicción de la intensidad del viento en altura (hasta 50 m) y que involucra la comparación del rendimiento de varios modelos fue descrito por Aydogmus et al. (2016), en el cual concluyen que, para la determinación de viabilidad de emplazamiento de parques eólicos, las máquinas de soporte vectorial representaron la mejor solución dentro de los siete modelos de regresión aplicados (p.179). Los modelos utilizados:

- Máquinas de soporte vectorial (*Support vector regression [SVMr]*)
- Perceptrón multicapa (*Multi-layer perceptron [MLP]*)
- Redes neuronales con función de base radial (*Radial basis function neural networks [RBFNetwork]*)
- KStar
- Aprendizaje ponderado localmente (*Locally weighted learning [LWL]*)
- Árboles de decisión (*DecisionStump*)
- Árboles aleatorios (*RandomTree*)

c. Métodos de ensamble

En términos generales, un método de ensamble se relaciona con la integración de dos o más modelos. Uno de los antecedentes de esta metodología es socializado por Ikuchi et al. (2018), el cual trata sobre el desarrollo de un algoritmo que integra las funciones de la predicción numérica del tiempo meteorológico y la predicción del viento

a corto tiempo (conocida en inglés como *nowcasting*) que se basa en datos de registrados por las aeronaves. Se utiliza el filtro de ensamble Kalman transformado para superar las limitaciones causadas por la ausencia de linealidad y normalidad (p. 365).

3.3.3 Usuarios de los datos de viento con fines aeronáuticos

a. Explotadores de aeronaves y miembros de la tripulación de vuelo

Estos usuarios requieren la información meteorológica para establecer políticas de seguridad operacional relacionadas con la planificación previa de las operaciones de vuelo, cambios en la planificación como medida de control operacional y evaluación de las condiciones atmosféricas antes de la salida y durante el vuelo en ruta de las aeronaves (OACI, 2019, p.5 – 1).

b. Dependencias de los servicios de tránsito aéreo (*Air Traffic Services [ATS]*)

A fin de que las dependencias ATS centralizadas logren una gestión eficiente de la afluencia del tránsito aéreo (*Air Traffic Flow Management [ATFM]*) y de la selección de derrotas, se ha hecho necesario para las computadoras del control de tránsito aéreo el ingreso de datos reticulares en forma digital de pronósticos del viento y de la temperatura en altitud a partir del sistema mundial de pronósticos de área (*World Area Forecast System [WAFS]*) (OACI, 2014, p.1-2).

4. MARCO METODOLÓGICO

4.1 Metodología

El estudio se basa principalmente en un método de investigación exploratorio-correlacional. Exploratorio, considerando que se presume como el primer estudio sobre el comportamiento de la intensidad del viento a diferentes alturas con aplicación particular en el diseño de rutas y procedimientos de vuelo. Correlacional, ante la necesidad de identificar la asociación que existe entre las variables meteorológicas involucradas para generar una herramienta de predicción de la intensidad del viento, partiendo de la medición individual de cada variable y posteriormente el establecimiento, cuantificación y análisis de sus relaciones.

De igual manera se considera la participación implícita del método experimental al identificar que, de acuerdo con la cantidad de observaciones disponibles, los modelos estadísticos diseñados parten de la división de los datos en dos grupos experimentales: 80% para entrenamiento y 20% para prueba. Con posterioridad a la selección de muestras representativas para cada función, los modelos realizan los experimentos correspondientes y miden la precisión de los resultados utilizando como métrica el error cuadrático medio (*Root Mean Square Error [RMSE]*), facilitando de esta manera el análisis de los modelos y la selección del más adecuado. El RMSE se define como:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2)$$

4.2 Enfoque

Como consecuencia de la naturaleza de los fenómenos meteorológicos analizados y el desarrollo de un método exploratorio-correlacional, durante el estudio se utilizan herramientas estadísticas que se relacionan con pruebas de hipótesis sobre la relación entre las variables y la generación de modelos capaces de aprender de esas relaciones

para generalizar el comportamiento de la intensidad del viento en condiciones particulares que son determinadas por la magnitud de las demás variables. Adicionalmente, frente al proceso deductivo y secuencial en el cual se basa el análisis objetivo de los resultados, se plantea la expectativa sobre una metodología que pueda ser replicada en otros estudios y la obtención de un modelo de predicción. Todas estas características conducen al reconocimiento de una investigación con enfoque cuantitativo.

4.3 Técnica: Herramientas para la obtención y tratamiento de datos

Se consolidó una base de datos para el periodo comprendido entre los años 2.015 y 2.020 a partir de la información climatológica registrada mediante lanzamientos de radiosonda desde la estación No. 80222 (Bogotá) y disponible en el sitio WEB del Departamento de Ciencia Atmosférica de la Universidad de Wyoming – Estados Unidos (<http://weather.uwyo.edu/upperair/sounding.html>). Se utilizó una hoja de cálculo de Microsoft Excel para la eliminación de registros con campos nulos y la generación de gráficas relacionadas con el comportamiento de las variables. La base de datos fue convertida en archivo con extensión .csv y almacenada en un repositorio de GITHUB (<https://raw.githubusercontent.com/LEXREYG/Statistics/master/bs1.csv>) para facilitar su lectura y compatibilidad con otras herramientas. El análisis exploratorio y descriptivo junto con el diseño de los modelos estadísticos se llevó a cabo en lenguaje Python mediante el uso de la plataforma online *Colaboratory*, la cual es un producto de *Google Research* diseñado para análisis de datos y desarrollo de tareas relacionadas con aprendizaje automático.

Tabla 1

Variables registradas mediante radiosonda

VARIABLE	DESCRIPCIÓN	UNIDADES
PRESION	Presión atmosférica	hPa
ALTURA	Altura geopotencial	m
TEMPAMB	Temperatura ambiente	°C
TEMPROCIO	Temperatura del punto de rocío	°C
HUMEDAD	Humedad relativa	%
MEZCLA	Relación de mezcla vapor/aire	g/k
VIENTO	Velocidad del viento	Nudos (<i>knots</i> [kt])
TPOT	Temperatura potencial	°K
TPOTEQ	Temperatura potencial equivalente	°K
TEPOTV	Temperatura potencial virtual	°K

Nota: La dirección del viento se omite por no ser una variable requerida en el diseño de rutas y procedimientos de vuelo, donde se considera que el viento es omnidireccional (OACI, 2020, p. I-2-3-6).

5. Diseño experimental

5.1 Análisis exploratorio y descriptivo

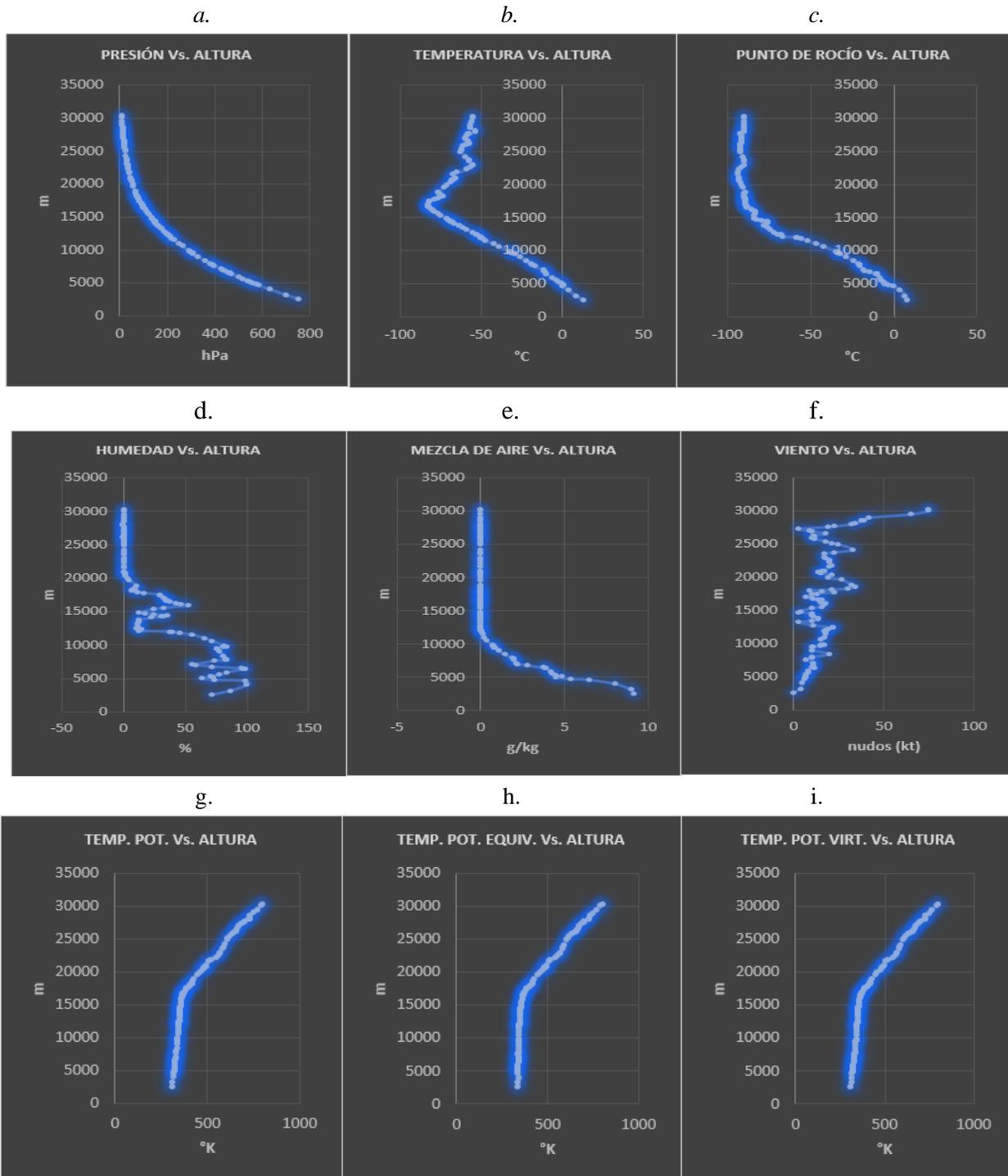
Este análisis se resume en el tratamiento de datos a partir de la identificación de las características de las variables y condiciones anómalas representadas en:

- Tipo de variable incorrecto (cualitativa, cuantitativa) y Valores atípicos
- Cruce entre variables numéricas y texto
- Valores ausentes y Caracteres especiales
- Colinealidad y Otras condiciones que el especialista considere de interés

Teniendo en cuenta que las aeronaves no vuelan a una altitud mayor de 50.000 ft en el espacio aéreo colombiano, se filtra la base de datos para conservar los registros con valores inferiores a 15.240 m. Para facilitar una familiarización general con las variables meteorológicas involucradas, la siguiente imagen representa gráficamente el comportamiento de cada una en función de la altura.

Imagen 3

Comportamiento de las variables en función de la altura



Nota. Las imágenes corresponden al registro de datos de radiosonda lanzada el 31 de diciembre de 2.020

5.1.1 Resumen general

a. Número de observaciones y valores ausentes

203.631 registros, sin campos nulos y 10 variables numéricas.

b. Resumen de los datos

Tabla 2

Estimadores estadísticos

	PRESION	ALTURA	TEMPAMB	TEMPROCIO	HUMEDAD	MEZCLAIRE	VIENTO	TPOT	TPOTEQ	TPOTV
count	203635.000000	203635.000000	203635.000000	203635.000000	203635.000000	203635.000000	203635.000000	203635.000000	203635.000000	203635.000000
mean	250.346213	13536.919989	-42.822289	-56.811329	33.496388	1.356126	19.004729	399.070314	403.572605	399.333385
std	213.522911	6881.566139	29.807513	34.871439	30.475899	2.569566	13.029474	100.321181	97.128121	100.117502
min	7.800000	2546.000000	-99.700000	-110.700000	0.000000	0.000000	0.000000	277.900000	278.400000	277.900000
25%	65.400000	7552.000000	-68.500000	-89.100000	4.000000	0.000000	10.000000	334.300000	341.600000	334.600000
50%	180.400000	13106.000000	-52.600000	-68.500000	26.000000	0.020000	16.000000	350.200000	350.500000	350.200000
75%	402.000000	19025.500000	-14.500000	-24.700000	57.000000	1.300000	24.000000	437.800000	437.800000	437.800000
max	760.000000	32558.000000	66.800000	28.800000	100.000000	49.880000	218.000000	944.800000	946.500000	944.900000

Nota. Las filas corresponden a los valores de los estimadores estadísticos en orden descendente así: cantidad de observaciones, media, desviación estándar, valor mínimo, primer cuartil, segundo cuartil, tercer cuartil y valor máximo.

5.1.2 Conclusiones del análisis univariado

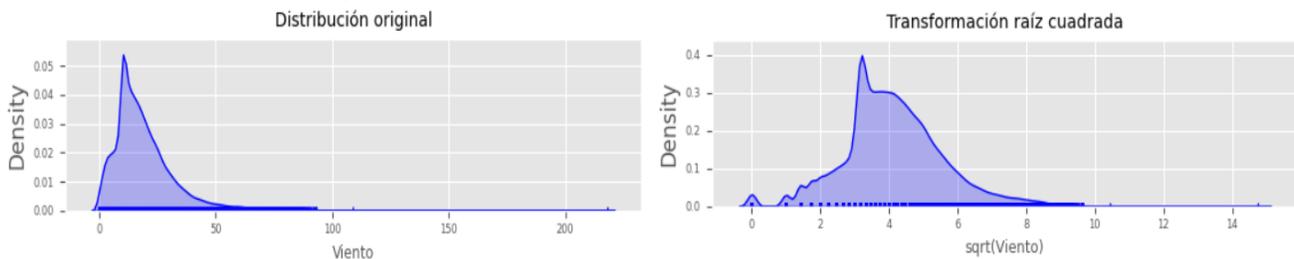
a. Identificación de valores atípicos

Se identificaron valores atípicos de temperatura mayores a 62° , razón de mezcla superiores a 30 g/kg e intensidad del viento mayor a 100 nudos, por lo cual se eliminan de la base de datos. Respecto a las demás variables, al considerar que son valores propios del comportamiento de la atmósfera, como por ejemplo la inversión térmica⁷ en la estratosfera (Imagen 3b. alturas mayores a 17.500 m), se decide continuar el proyecto sin eliminar los valores atípicos identificados.

b. Distribución de la variable objetivo: VIENTO

Imagen 4

Distribución de la variable objetivo

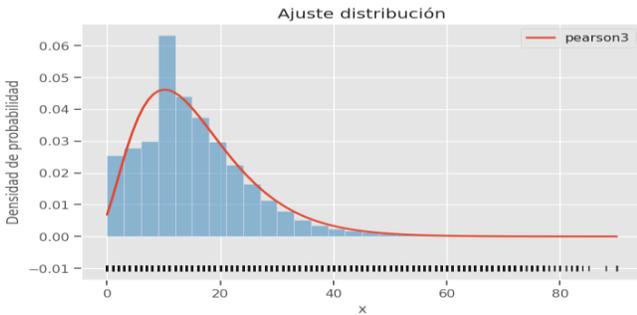


⁷ En la troposfera (capa en contacto con la superficie terrestre) la temperatura disminuye con la altura. Este comportamiento se invierte en la estratosfera (capa superior a la troposfera).

Debido a la forma de los datos, resulta útil identificar la distribución que se ajusta mejor a los datos y a partir de la cual se identifican los parámetros que con mayor probabilidad generan los valores registrados:

Imagen 5

Ajuste de la distribución pearson3



Nota: Los resultados del ajuste son: Dominio: [-inf, inf], Parámetros: {'skew': 1.1016559153628296, 'loc': 15.692049015884304, 'scale': 9.990014641869813}, Log likelihood: -420538.8770835993, AIC: 841083.7541671986, BIC: 841112.7496644477

Donde skew, loc y scale corresponden con el sesgo, la media y desviación estándar respectivamente.

c. Variables predictoras

A partir del análisis de distribución de las variables predictoras, se identificó que no hay presencia de sesgo significativo en la distribución de frecuencias, excepto por la variable MEZCLAIRE; no obstante, bajo la hipótesis de tratar con variables altamente aleatorias, el análisis y procedimientos posteriores se realizaron sin intervenir la geometría observada en procura de evitar su alteración.

5.1.3 Conclusiones del análisis bivariado

Se observó una aparente correlación de las variables predictoras con la variable objetivo, sin que fuera posible determinar la existencia de una condición lineal. Por esta razón se realiza genera la matriz de correlación de Pearson, con el fin de identificar numéricamente el nivel de correlación, donde los resultados permiten concluir, según la Imagen 6, que la variable VIENTO presenta los valores más bajos de correlación respecto a las demás variables, con valores mínimo y máximo 0.26 (con TPOTEQ) y 0.41 (con TEMPROCIO) respectivamente.

Se evidencia también posibles problemas de colinealidad a causa de la correlación entre las variables TPOT, TPOTEQ y TPOTV, por esta razón se integran en una sola variable (TPOTP) a partir de su promedio y se revisa la matriz de correlación (Imagen 7).

Imagen 6

Matriz de correlación inicial

PRESION	1	-0.98	0.96	0.95	0.57	0.9	-0.39	-0.99	-0.73	-0.99
ALTURA	-0.98	1	-0.99	-0.97	-0.55	-0.83	0.4	0.98	0.8	0.98
TEMPAMB	0.96	-0.99	1	0.96	0.51	0.78	-0.39	-0.95	-0.82	-0.95
TEMPROCIO	0.95	-0.97	0.96	1	0.71	0.85	-0.41	-0.96	-0.71	-0.96
HUMEDAD	0.57	-0.55	0.51	0.71	1	0.72	-0.29	-0.61	-0.2	-0.61
MEZCLAIRE	0.9	-0.83	0.78	0.85	0.72	1	-0.36	-0.89	-0.39	-0.88
VIENTO	-0.39	0.4	-0.39	-0.41	-0.29	-0.36	1	0.38	0.26	0.38
TPOT	-0.99	0.98	-0.95	-0.96	-0.61	-0.89	0.38	1	0.76	1
TPOTEQ	-0.73	0.8	-0.82	-0.71	-0.2	-0.39	0.26	0.76	1	0.77
TPOTV	-0.99	0.98	-0.95	-0.96	-0.61	-0.88	0.38	1	0.77	1

Imagen 7

Matriz de correlación revisada

PRESION	1	-0.98	0.96	0.95	0.57	0.9	-0.98	-0.39
ALTURA	-0.98	1	-0.99	-0.97	-0.55	-0.83	0.98	0.4
TEMPAMB	0.96	-0.99	1	0.96	0.51	0.78	-0.96	-0.39
TEMPROCIO	0.95	-0.97	0.96	1	0.71	0.85	-0.94	-0.41
HUMEDAD	0.57	-0.55	0.51	0.71	1	0.72	-0.55	-0.29
MEZCLAIRE	0.9	-0.83	0.78	0.85	0.72	1	-0.82	-0.36
TPOTP	-0.98	0.98	-0.96	-0.94	-0.55	-0.82	1	0.37
VIENTO	-0.39	0.4	-0.39	-0.41	-0.29	-0.36	0.37	1

Tabla 3*Estimadores estadísticos definitivos*

	PRESION	ALTURA	TEMPAMB	TEMPROCIO	HUMEDAD	MEZCLAIRE	TPOTP	VIENTO
count	116448.000000	116448.000000	116448.000000	116448.000000	116448.000000	116448.000000	116448.000000	116448.000000
mean	392.630699	8394.604098	-23.652319	-32.483613	51.210516	2.368676	337.966319	15.692053
std	178.271018	3565.085333	24.924824	27.123187	27.716068	3.019120	10.451352	10.089591
min	124.600000	2546.000000	-88.100000	-105.100000	0.010000	0.000100	301.433333	0.010000
25%	243.000000	5432.000000	-42.300000	-55.200000	28.000000	0.090000	329.000000	10.000000
50%	376.000000	8065.000000	-18.500000	-31.200000	51.000000	0.790000	338.400000	14.000000
75%	529.000000	11180.000000	-2.500000	-8.500000	74.000000	3.930000	346.400000	21.000000
max	760.000000	15240.000000	24.600000	15.600000	100.000000	15.030000	360.033333	90.000000

Nota. Las filas corresponden a los valores de los estimadores estadísticos en orden descendente así: cantidad de observaciones, media, desviación estándar, valor mínimo, primer cuartil, segundo cuartil, tercer cuartil y valor máximo.

5.2 Modelamiento

5.2.1 Procedimiento

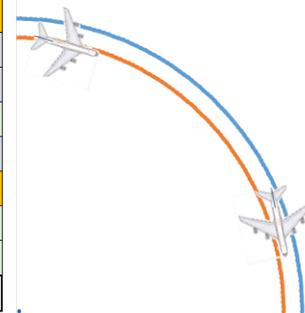
Amat (2020), recomienda una serie de pasos para el diseño de modelos de aprendizaje automático orientado a la identificación de patrones como insumo para la generación de estructuras que representan el comportamiento de los fenómenos. En consecuencia, los modelos incluidos en este estudio observan el siguiente procedimiento:

- Definición del problema:** Se pretende garantizar a los diseñadores de rutas y procedimientos de vuelo la disponibilidad de datos sobre la intensidad del viento en el espacio aéreo de Bogotá, en intervalos de 30,48 m (100 ft) desde una altitud de 2.560 m (8.400 ft) hasta 9.144 m (30.000 ft). El análisis considera la premisa de utilizar datos que representen las condiciones más críticas, debido a que las áreas de protección trazadas observan la probabilidad para la aeronave de encontrar condiciones de baja presión atmosférica, altas temperaturas y viento omnidireccional. De igual manera, es necesario advertir que la información obtenida del modelo debe permitir el establecimiento de un valor de intensidad del viento con una probabilidad de ocurrencia del 95% (2σ).

Imagen 8

Ejemplo del efecto de la altitud, inclinación lateral de la aeronave, presión, temperatura e intensidad del viento en la velocidad y desplazamiento de una aeronave

INPUT		INPUT	
LEVEL (FEET)	17000	LEVEL (FEET)	19000
IAS (KT)	220	IAS (KT)	220
w (ICAO) (KT)	81	w (ICAO) (KT)	85
BANK (°)	15	BANK (°)	15
OUTPUT		OUTPUT	
TAS (KT)	294.8	TAS (KT)	303.6
GS (KT)	375.8	GS (KT)	388.6
r (NM)	7.68	r (NM)	8.21



Nota. Una aeronave de iguales características, a una misma velocidad IAS⁸ y a diferentes niveles, describe trayectorias de vuelo diferentes debido al efecto de la altitud y el viento en la determinación del radio de viraje (r). Se indica en las tablas de datos el resultado de las correcciones de velocidad, obteniendo como resultado la TAS⁹ y GS¹⁰. La aplicación del viento OACI indicada en la ecuación (1) en función de la altitud que genera en el primer caso un radio de viraje de 7,68 NM (trayectoria naranja a 17.000 ft) y de 8,21 NM en el segundo caso (trayectoria azul a 19.000 ft).

5.2.2 Métrica

La metodología para seleccionar el mejor modelo se basó en la evaluación de cuatro aspectos:

- (1) Valor RMSE, el cual es una métrica común en modelos de regresión.
- (2) Margen de error de una estimación con relación a una misma observación seleccionada aleatoriamente.
- (3) Diagrama de la dispersión de los individuos Predicciones Vs. Observaciones, respecto a la pendiente de referencia del modelo.
- (4) Distribución de la probabilidad de valores de error de las predicciones.

5.2.3 Estrategia de evaluación del modelo

Con el fin de facilitar la evaluación de los modelos con relación al error de predicción, se divide la base de datos en dos conjuntos: entrenamiento y prueba. Debido a la cantidad disponible de observaciones, el porcentaje asignado a cada conjunto fue de 80% y 20% respectivamente, acompañado de la verificación de similitud entre los estimadores estadísticos del conjunto de entrenamiento y prueba.

5.2.4 Preprocesamiento de los datos

Se aplicaron transformaciones de normalización mediante la ecuación (3), a los datos cuantitativos como medio para facilitar su interpretación por parte de los algoritmos de aprendizaje automático desarrollados.

$$z = \frac{x - \mu}{\sigma} \quad (3)$$

El procesamiento se complementó con las herramientas de entrenamiento y transformación integradas en la función *pipeline* y *column transformer*, para aprendizaje previo de parámetros mediante las observaciones de entrenamiento y posterior transformación múltiple y aplicación a los datos.

5.2.5 Ajuste de un modelo base

Se modeló un primer algoritmo de regresión lineal, el cual se utilizó para evaluar su comportamiento frente a los datos, se optimizó el modelo mediante la identificación de los mejores hiperparámetros como preparación para la prueba de predicción de valores de viento ausentes en la base de datos.

⁸ IAS: *Indicated Air Speed* o velocidad indicada la cual se basa en la lectura directa del anemómetro de la aeronave.

⁹ TAS: *True Air Speed* o velocidad indicada verdadera. Corresponde a la IAS corregida por altitud, temperatura y presión.

¹⁰ GS: *Ground Speed* o velocidad respecto al suelo corregida por viento.

5.2.6 Algoritmos de machine learning

a) Proceso general

- Estandarización de variables predictoras
- Integración de transformadores y entrenamientos en un pipeline
- Entrenamiento, predicción y determinación del error mediante conjuntos de prueba
- Validación de las observaciones no conocidas mediante repetición con múltiples conjuntos de entrenamiento, donde el error estimado se obtiene de todas las estimaciones individuales.
- Ajuste automático de hiperparámetros

b) Algoritmos diseñados

Se diseñó un total de 10 algoritmos de acuerdo con la cantidad y naturaleza indicados a continuación:

- 1 Modelo *RIDGE*
- 1 Modelo *K-Nearest Neighbor* (kNN)
- 1 Regresión lineal (Ridge y Lasso)
- 2 *Random Forest*
- 2 *Gradient Boosting Trees*
- 1 *Stacking*
- 2 Red Neuronal

6. Resultados

Con posterioridad al ajuste de parámetros con la intención de encontrar, inclusive entre modelos de la misma categoría, el de mejor rendimiento, se procedió a generar los grupos de entrenamiento y prueba a partir de la base de datos, la integración de las funciones de estandarización y transformación, los bloques de predicción y los elementos de evaluación, obteniendo los siguientes resultados para una observación aleatoria cuyos valores de las variables se indican a continuación:

PRESION	460	ALTURA	6549	TEMPAMB	-9.40	TEMPROCIO	-19.40
HUMEDAD	44	MEZCLAIRE	1.80	TPOTP	331.56	VIENTO	10

De acuerdo con lo explicado en la métrica de evaluación, los cuatro criterios se encuentran plasmados en las Tablas 4 a 8, las cuales contienen los pares de modelos por categoría sobre los cuales se realizó la comparación y evaluación. Como conclusión, el modelo *Gradient Boosting 1* presentó las mejores condiciones en cada uno de los criterios:

- (1) Valor más bajo en la métrica RMSE: **8,78 nudos**
- (2) Error en la predicción: **3,37 nudos**
- (3) Gráfica de dispersión: **Menor dispersión asociada al RMSE**
- (4) Distribución del error: **Valor mayor absoluto de probabilidad de error: 38** (cercano al valor indicado por el modelo k-NN). Los demás reflejan un valor cercano a 60.

6.1 Discusión de los resultados

Con el fin de identificar las ventajas y la aplicación práctica de los resultados obtenidos mediante el modelo seleccionado, se realizó la predicción del valor de intensidad del viento a partir de los parámetros de 10 observaciones seleccionadas aleatoriamente de la base de datos que contiene registros entre 2.546 m y 15240 m.

La evaluación se realiza con los valores de altura, intensidad del viento observado, la predicción de la intensidad del viento hecha por el modelo, el intervalo de valores de viento a la altura de la observación, el número de observaciones existentes para la misma altura y el valor del viento OACI para cada nivel. Los resultados se resumen en la Tabla 9.

El análisis de los resultados desde la óptica de la gestión del tránsito aéreo se resume en las observaciones siguientes:

- La probabilidad de ocurrencia de viento con valores iguales o superiores al viento OACI es prácticamente 0%.
- Existen predicciones con un valor menor al observado (Muestras 1, 5, 8 y 10). Teniendo en cuenta que según se explica en la definición del problema suministrada en 4.4.2 e Imagen 8, esta condición no es deseable porque conduce al diseño de áreas con un valor de radio de viraje inferior, motivando la omisión de obstáculos que deberían ser incluidos en el análisis de las rutas o procedimientos.
- Existen predicciones con un valor igual o muy cercano al valor de la observación (Muestra 2) y predicciones con valores superiores a las observaciones (Muestras 3, 4, 6, 7 y 9). Ambos casos generan la falsa percepción de estar utilizando valores de radio iguales o mayores en el diseño de las áreas de protección en función del viento característico a ese nivel particular.

Tabla 4

Comparación de resultados para modelos basados en Gradient Boosting Trees

	GRADIENTBOOSTING 1	GRADIENTBOOSTING2 2
RMSE	8,78	8,86
PRED	13,37	13,61

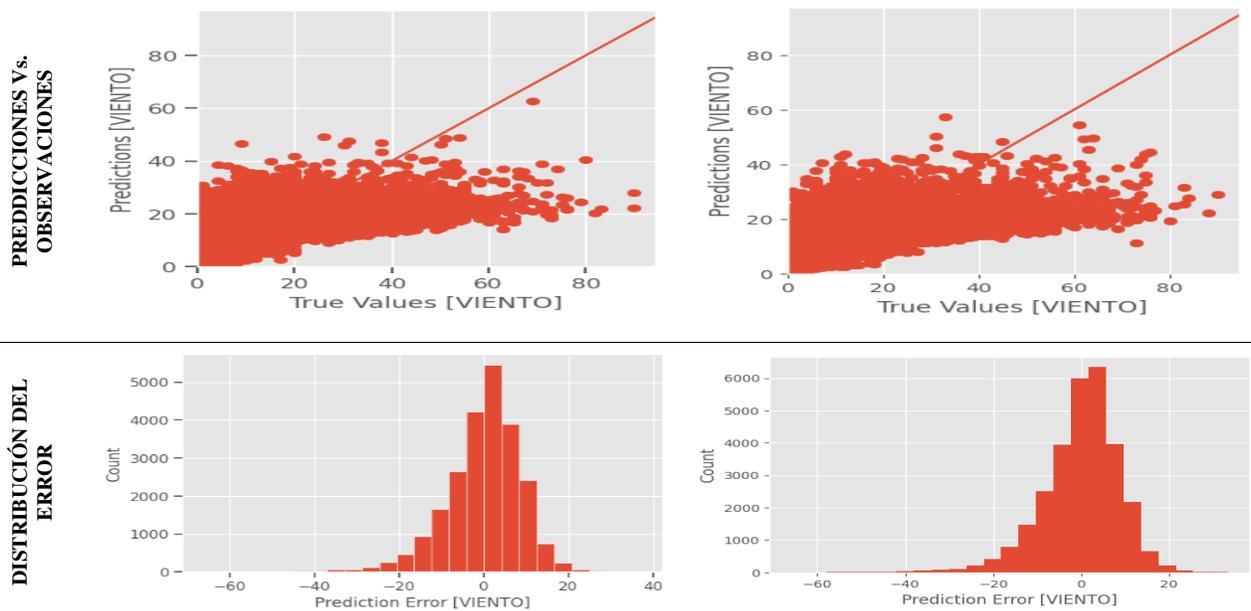


Tabla 5

Comparación de resultados para modelos basados en Redes Neuronales

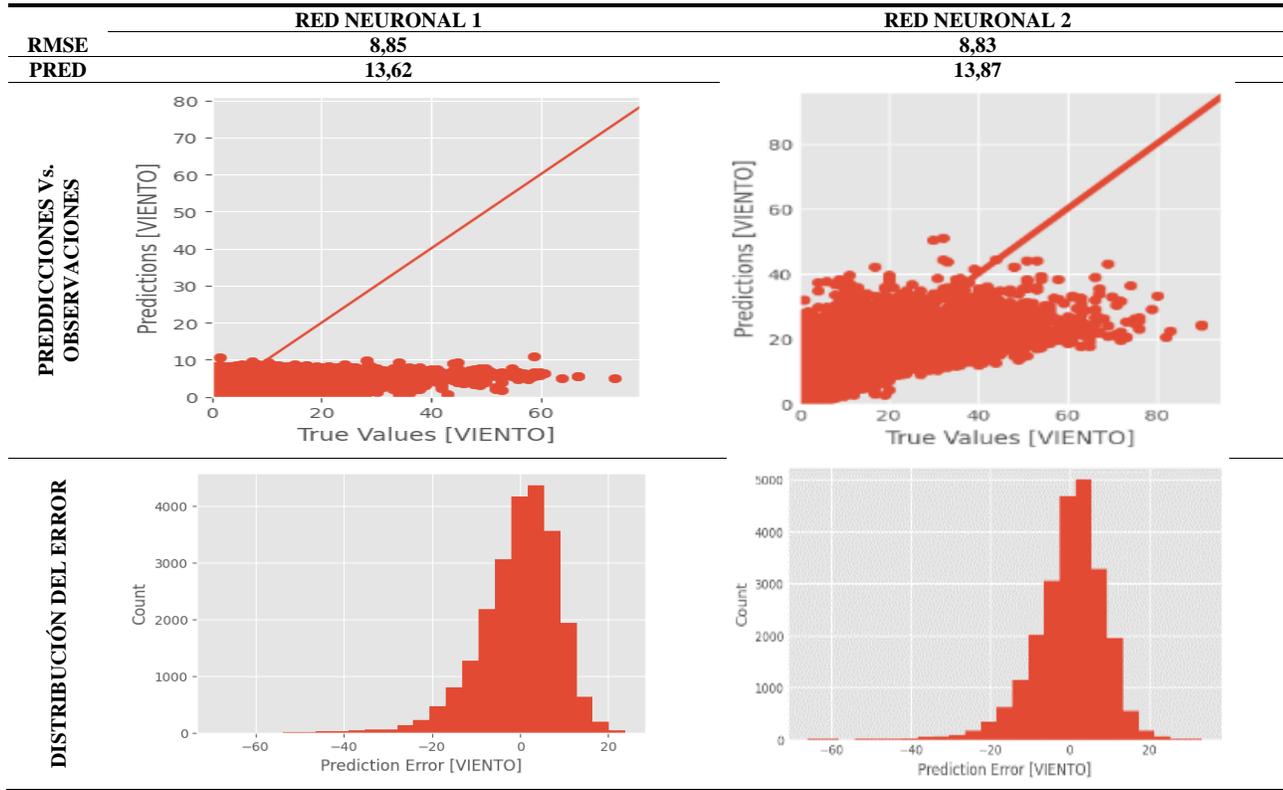


Tabla 6

Comparación de resultados para modelos basados en Métodos de ensamble y evaluación de conjuntos de datos

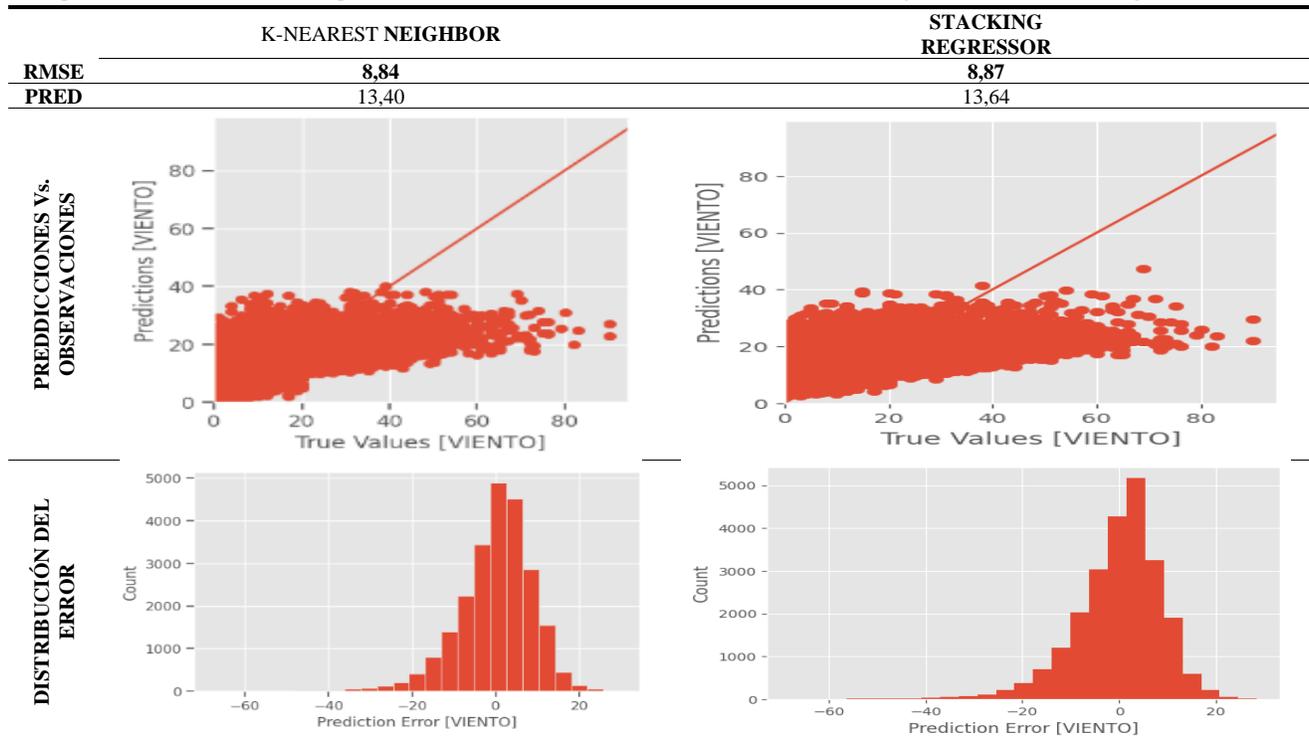


Tabla 7

Comparación de resultados para modelos basados en Árboles de decisión

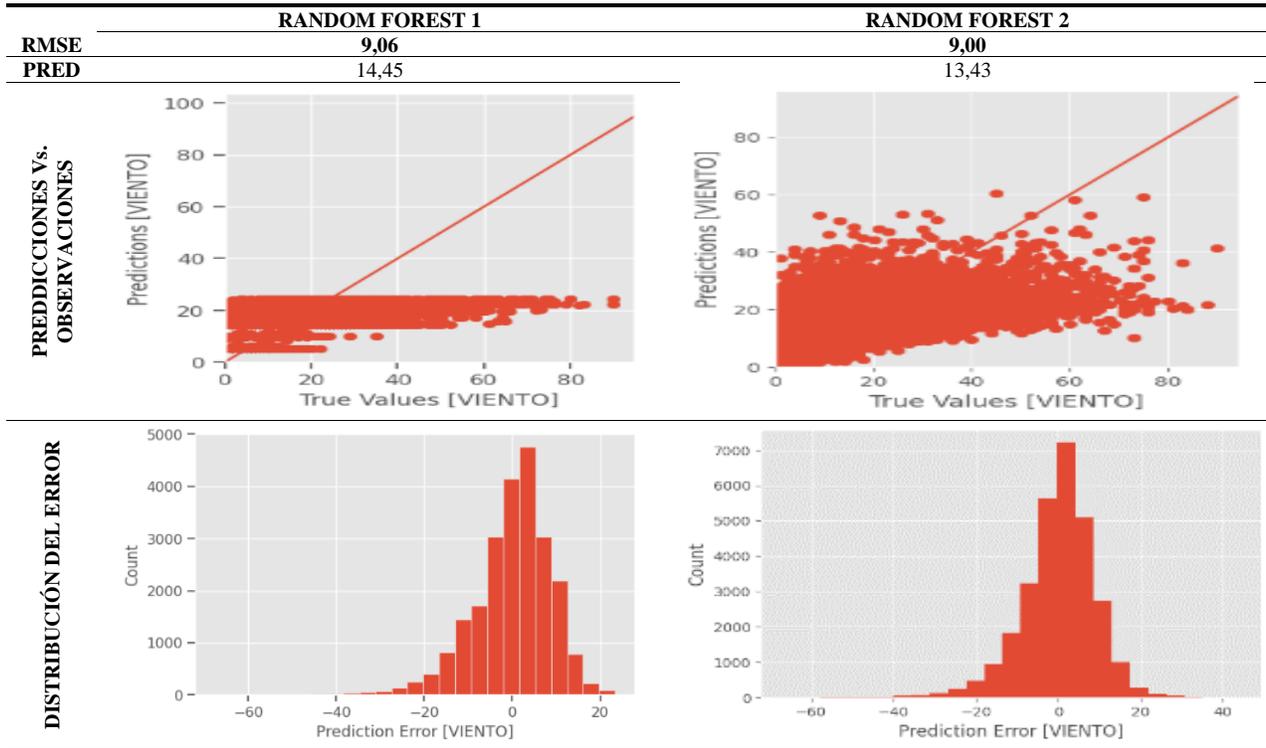


Tabla 8

Comparación de resultados para modelos basados en Modelos Lineales y de regularización

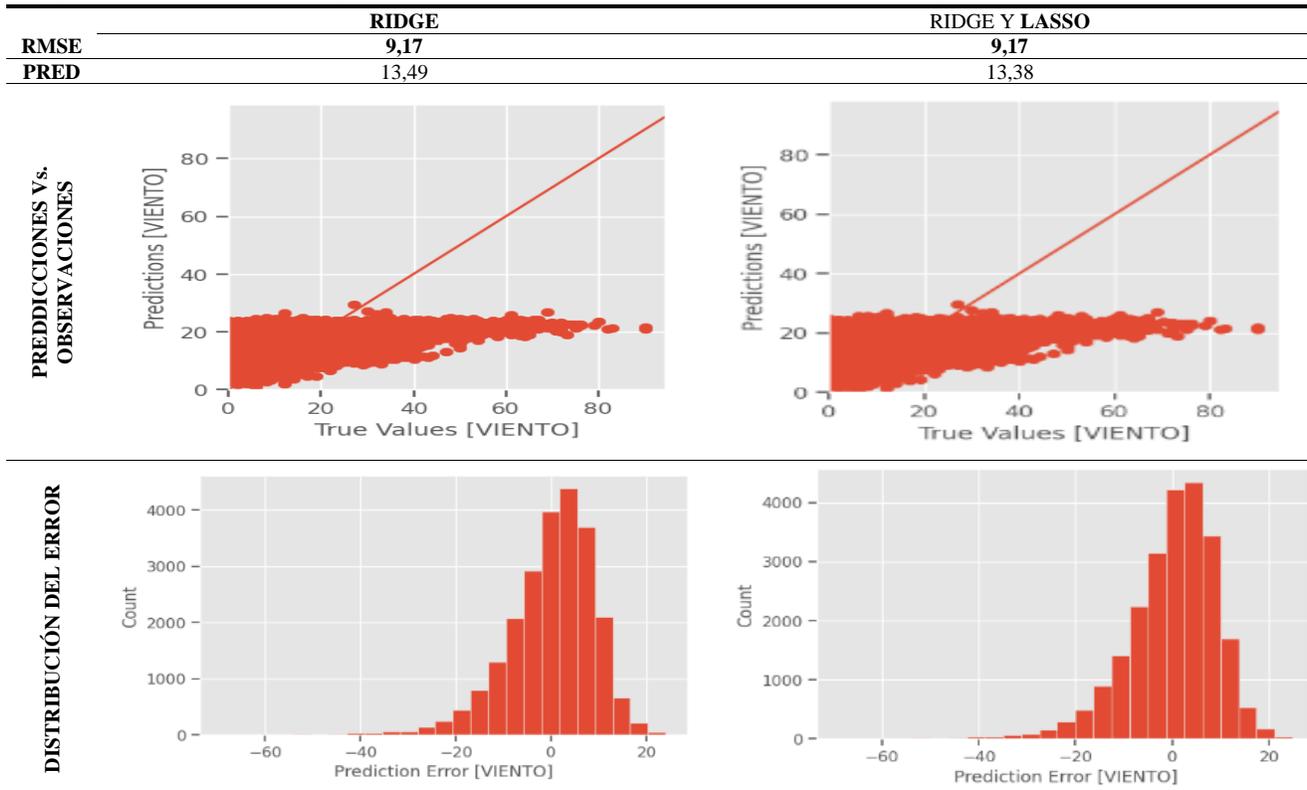


Tabla 9

Resultados de las predicciones a partir de observaciones seleccionadas aleatoriamente

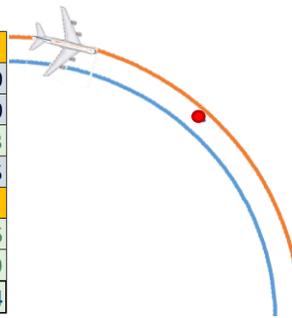
MUESTRA	ALTURA (m - ft)	No. OBSERVACIONES MISMA ALTURA	OBSERVACIÓN PROBABILIDAD (nudos - %)	PREDICCIÓN PROBABILIDAD (nudos - %)	INTERVALO (nudos)	VIENTO OACI PROBABILIDAD (nudos - %)
1	2.929 – 9.600	7	5 – 85,53	2,83 – 89,88	1 - 11	66,21 – $2,7 \times 10^{-5}$
2	2.569 – 8.400	193	6 – 83,16	6,03 – 83,08	0 - 12	63,85 – $9,03 \times 10^{-5}$
3	4.514 – 14.800	7	5 – 85,53	14 – 56,65	3 - 28	76,61 – $7,76 \times 10^{-8}$
4	5.632 – 18.500	11	10 – 71,36	14,25 – 56,68	10 - 26	83,95 – $6,6 \times 10^{-10}$
5	3.311 – 10.800	6	17 – 44,83	9,73 – 72,27	4 - 17	68,72 – $7,33 \times 10^{-6}$
6	3.962 – 13.000	582	10 – 71,36	16,68 – 45,30	0 - 37	73 – $6,7 \times 10^{-7}$
7	4.993 – 16.400	10	9 – 74,64	14,10 – 56,26	4 - 18	79,76 – $1,07 \times 10^{-8}$
8	5.923 – 19.400	26	28 – 11,12	13,73 – 57,70	1 - 36	85,86 – $1,75 \times 10^{-10}$
9	6.976 – 22.900	10	3 – 89,58	13,32 – 59,29	3 - 21	91,77 – $2,32 \times 10^{-12}$
10	11.773 – 38.700	8	22 – 26,58	17,84 – 41,56	2 - 45	124,25 – $2,62 \times 10^{-25}$

Nota. La probabilidad se relaciona con valores mayores o iguales a los indicados. El intervalo indica el valor mínimo y máximo de todas las observaciones registradas para esa altura.

Imagen 9

Radios de viraje basados en los resultados de la **Muestra 8 - Tabla 9**

INPUT		INPUT	
LEVEL (FEET)	19400	LEVEL (FEET)	19400
IAS (KT)	220	IAS (KT)	220
OBSERVED WIND	28	PREDICTED WIND	13.73
BANK (°)	15	BANK (°)	15
OUTPUT		OUTPUT	
TAS (KT)	305.4	TAS (KT)	305.36
GS (KT)	333.4	GS (KT)	319.09
r (NM)	6.04	r (NM)	5.54

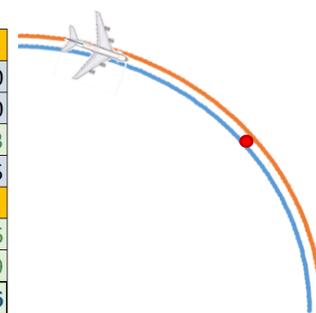


Nota. El punto entre los arcos de la gráfica de la derecha representa un obstáculo. El arco azul representa el uso de un radio de viraje calculado a partir de un viento de 13,73 nudos, con el cual se omite el obstáculo indicado. El arco naranja representa el uso de un radio de viraje calculado a partir de un viento de 28 nudos (trayectoria más probable de la aeronave), con el cual se incluye el obstáculo.

Imagen 10

Radios de viraje basados en los resultados de la **Muestra 2 - Tabla 9**

INPUT		INPUT	
LEVEL (FEET)	8400	LEVEL (FEET)	8400
IAS (KT)	220	IAS (KT)	220
OBSERVED WIND	12	PREDICTED WIND	6.03
BANK (°)	15	BANK (°)	15
OUTPUT		OUTPUT	
TAS (KT)	257	TAS (KT)	256.96
GS (KT)	269	GS (KT)	262.99
r (NM)	3.93	r (NM)	3.76



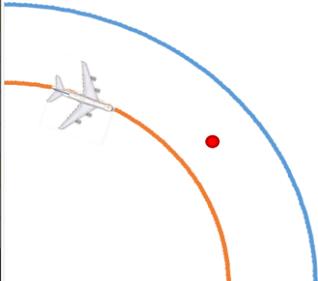
Nota. El punto entre los arcos de la gráfica de la derecha representa un obstáculo. El arco azul representa el uso de un radio de viraje calculado a partir de un viento de 6,03 nudos, con el cual se omite el obstáculo indicado. El arco naranja representa el uso de un radio de viraje calculado a partir de un viento de 12 nudos (trayectoria a base del viento máximo registrado para esa altura), con el cual se incluye el obstáculo.

- El valor de viento OACI se encuentra por fuera y muy lejos del intervalo observado para cada altura, lo que, contrario a lo expresado en los ítems anteriores, genera que el diseño de áreas de protección con este viento tenga lugar con valores de radio de viraje exagerados, causando la inclusión de obstáculos que eventualmente causan restricciones innecesarias para las aeronaves, afectando la eficiencia y la gestión del espacio aéreo.

Imagen 11

Radios de viraje basados en los resultados de la **Muestra 4 - Tabla 9**

INPUT		INPUT	
LEVEL (FEET)	18500	LEVEL (FEET)	18500
IAS (KT)	220	IAS (KT)	220
OBSERVED WIND	26	ICAO w	83.95
BANK (°)	15	BANK (°)	15
OUTPUT		OUTPUT	
TAS (KT)	301.4	TAS (KT)	301.4
GS (KT)	327.4	GS (KT)	385.35
r (NM)	5.83	r (NM)	8.08



Nota. El punto entre los arcos de la gráfica de la derecha representa un obstáculo. El arco azul representa el uso de un radio de viraje calculado a partir del viento OACI de 83,95 nudos, con el cual se incluye innecesariamente el obstáculo indicado, causando eventualmente restricciones operacionales para la aeronave o de gestión del espacio aéreo. El arco naranja representa el uso de un radio de viraje calculado a partir de un viento de 26 nudos (trayectoria más probable), con el cual se omite el obstáculo.

Lo expuesto previamente se justifica en la existencia de múltiples observaciones para un mismo valor de altura (Tabla 9, columna 3), lo que genera un intervalo de valores de viento del cual se derivan valores mínimo y máximo de viento (Tabla 9, columna 5). Esto permite concluir que es necesario identificar la posición del valor de la predicción dentro del intervalo existente, y de esta manera determinar la utilidad del dato obtenido mediante el modelo.

Con el análisis de los resultados iniciales, se identificaron valores cercanos a los 15 nudos (por la influencia de los dos valores de la moda: 0 nudos y 10 nudos) en su mayoría inferiores a los del viento registrado por la radiosonda, condición que no favorece la disponibilidad de datos de viento utilizables para el diseño de rutas y procedimientos de vuelo; teniendo en cuenta que son las condiciones más críticas las que deben ser consideradas. De igual manera se evidenció que el viento OACI en ningún caso corresponde con los valores de las predicciones u observaciones, al presentar casos que superan en 45 nudos a los registrados por la radiosonda.

Lo anterior conduce a la necesidad de evaluar la manera en que se preparan los datos para ser sometidos al modelo y de esta manera obtener una herramienta útil para la gestión del tránsito aéreo. El procedimiento observado para tal efecto se indica en los siguientes pasos:

- Mediante una tabla dinámica se generó una nueva base de datos con registros individuales para cada valor de altura suministrado por la radiosonda, asignando a cada registro el valor mínimo de presión atmosférica, humedad y razón de mezcla de aire, y valor máximo para cada uno de los tipos de temperatura.
- Se aplicó el modelo a los datos resultantes, obteniendo un RMSE de 9,24 nudos.
- La base de datos resultante cuenta con 11.334 filas y 8 variables y se encuentra disponible en Github para acceso a través de Google Colab: <https://raw.githubusercontent.com/LEXREYG/Statistics/master/bsmax.csv>. Los estadísticos de esta nueva base de datos se indican en la siguiente tabla.

Tabla 10*Estimadores estadísticos para la base de datos ajustada*

	PRESION	ALTURA	TEMPAMB	TEMPROCIO	HUMEDAD	MEZCLAIRE	TPOTP	VIENTO
count	11334.000000	11334.000000	11334.000000	11334.000000	11334.000000	11334.000000	11334.000000	11334.000000
mean	379.398077	8636.070496	-24.075419	-29.360085	26.019587	1.327410	340.490995	26.876654
std	174.882158	3538.885600	24.983001	26.488538	23.277576	2.306465	10.003254	12.376843
min	126.000000	2546.000000	-76.500000	-85.100000	0.000000	0.000000	312.833333	0.000000
25%	225.000000	5523.250000	-45.500000	-52.100000	8.000000	0.030000	332.100000	19.000000
50%	348.000000	8614.500000	-21.100000	-26.800000	18.000000	0.170000	341.766667	25.000000
75%	522.000000	11669.750000	-1.100000	-5.300000	39.000000	1.350000	348.500000	33.000000
max	750.000000	15234.000000	24.600000	15.600000	100.000000	12.160000	360.033333	90.000000

A partir de los nuevos datos se procedió con la identificación del comportamiento de las predicciones respecto a las observaciones y los valores de viento OACI, se analizó nuevamente el intervalo de niveles entre 5.182m (17.000 ft) y 5.791 m (19.000 ft). En esta oportunidad se observó que las predicciones mantuvieron valores medios con relación a los valores observados (aproximadamente 25 nudos coincidente con el valor de la moda).

Los únicos valores de interés para el diseño de procedimientos de vuelo son aquellos de viento observado mayores a las predicciones del modelo, ya que representan la condición más crítica del viento y son la base para la toma de decisiones con relación al establecimiento de criterios para la utilización de los datos. Al analizar el conjunto de datos representado en la imagen anterior, se identifica una diferencia máxima entre la predicción y la observación (valor máximo) de 15 nudos; 2,7 nudos más que la desviación estándar. Adicionalmente, en apartados previos se identificó que los valores de viento OACI son improbables, con lo cual se plantea la posibilidad de sumar a cada predicción el valor de 15 nudos, con lo cual se obtendría el valor máximo posible para el rango de niveles evaluado; lo cual resulta altamente significativo si se observa que, con relación al viento OACI, se logra una reducción del viento entre 40 y 45 nudos.

Tomando como referencia el nivel más alto (5.683 M – 18.400 ft), la diferencia en el valor de radio calculado a partir de la predicción sumando 15 nudos y el viento OACI, se muestra en la Imagen 12. Una vez propuesto el método para la utilización del modelo, se procedió con la predicción de los valores de viento para cada nivel de interés dentro del espacio aéreo de Bogotá, según se indicó en la definición del problema en el numeral 4.4.2. El modelo obtenido requiere el ingreso de los valores de cada variable, lo que sugiere la imputación ante la ausencia de estos. El método de imputación consistió en ubicar el valor de interés en medio de dos valores observados, de manera que los tres valores resultaran ordenados de menor a mayor. Posteriormente se imputa el valor de cada variable a partir del promedio entre el valor anterior y el valor posterior, obteniendo un valor de referencia para la intensidad del viento con el cual fuera posible evaluar las predicciones.

Imagen 12*Comparación de valores de radio*

INPUT		INPUT	
LEVEL (FEET)	18644	LEVEL (FEET)	18644
IAS (KT)	220	IAS (KT)	220
PREDICTED WIND	43	ICAO w	84.288
BANK (°)	15	BANK (°)	15
OUTPUT		OUTPUT	
TAS (KT)	302	TAS (KT)	302.03
GS (KT)	345	GS (KT)	386.32
r (NM)	6.47	r (NM)	8.12



Nota. El arco azul representa el uso de un radio de viraje calculado a partir del viento OACI de 84,288 nudos. El arco naranja representa el uso de un radio de viraje calculado a partir de un viento de una predicción de 43 nudos. La diferencia en el radio de viraje de la aeronave es de 1,65 NM (3,0 km).

Los resultados de las predicciones en función de los valores imputados se muestran en la Imagen 13. Se observa que el modelo sigue el patrón de los datos de referencia, presentando una diferencia máxima positiva de 12,41 nudos, valor cercano a la desviación estándar de los datos. Al aplicar la metodología de sumar el valor máximo de error a todas las predicciones, se obtienen 217 valores de viento para ser aplicados en el diseño de procedimientos de vuelo desde 2.560,35 m (8.400 ft) hasta 9.144,12 m (30.000 ft).

El nuevo conjunto de observaciones de variables predictoras que acompañan a la intensidad del viento presenta las características indicadas en la Tabla 12, donde por una media de 34,47 y desviación estándar de 5,21, el valor del viento que representa en cualquier caso una probabilidad del 95% es 43,05 nudos.

De acuerdo con el criterio del diseñador de procedimientos y del escenario en particular, es posible utilizar los valores individuales de la intensidad de viento en cada nivel o el valor correspondiente al 95% de probabilidad para el trazado de áreas de protección a partir del cálculo del radio de viraje.

Tabla 11

Comparación de resultados: Base de datos inicial base de datos con imputación de variables

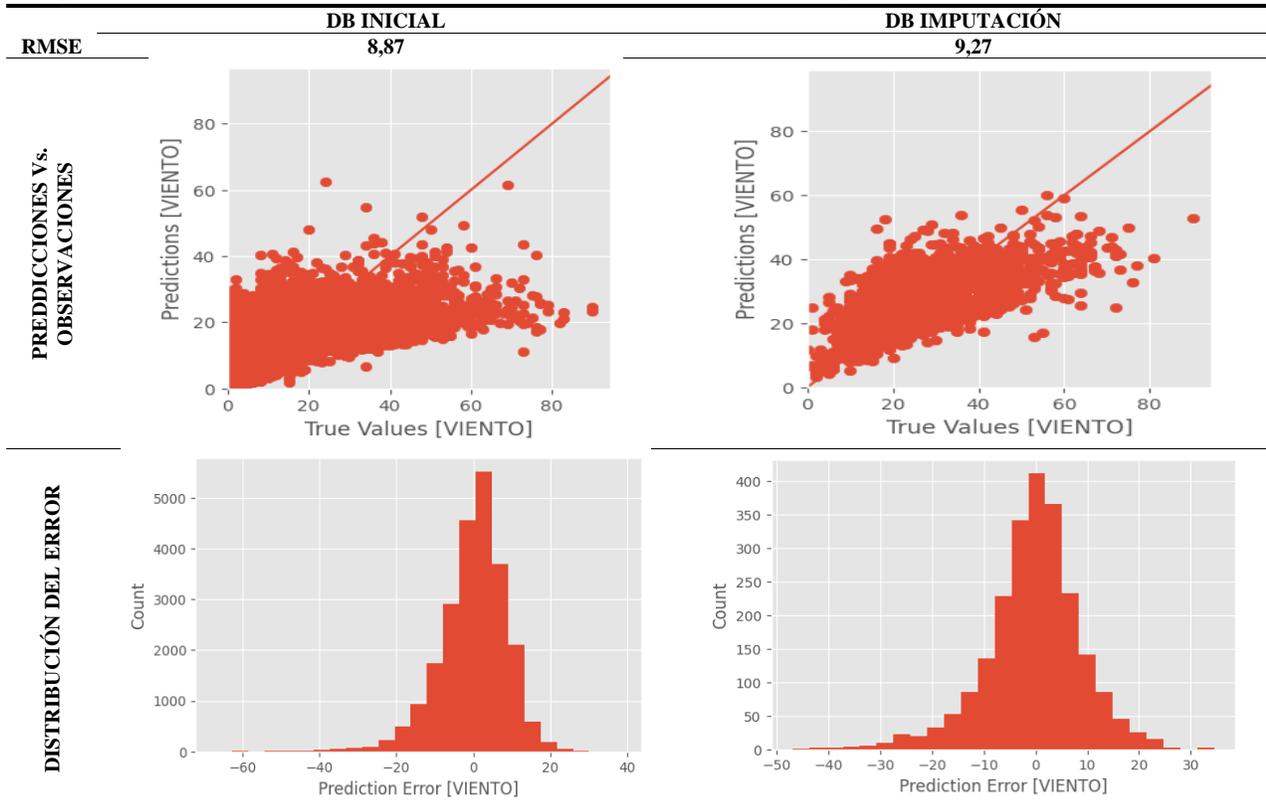


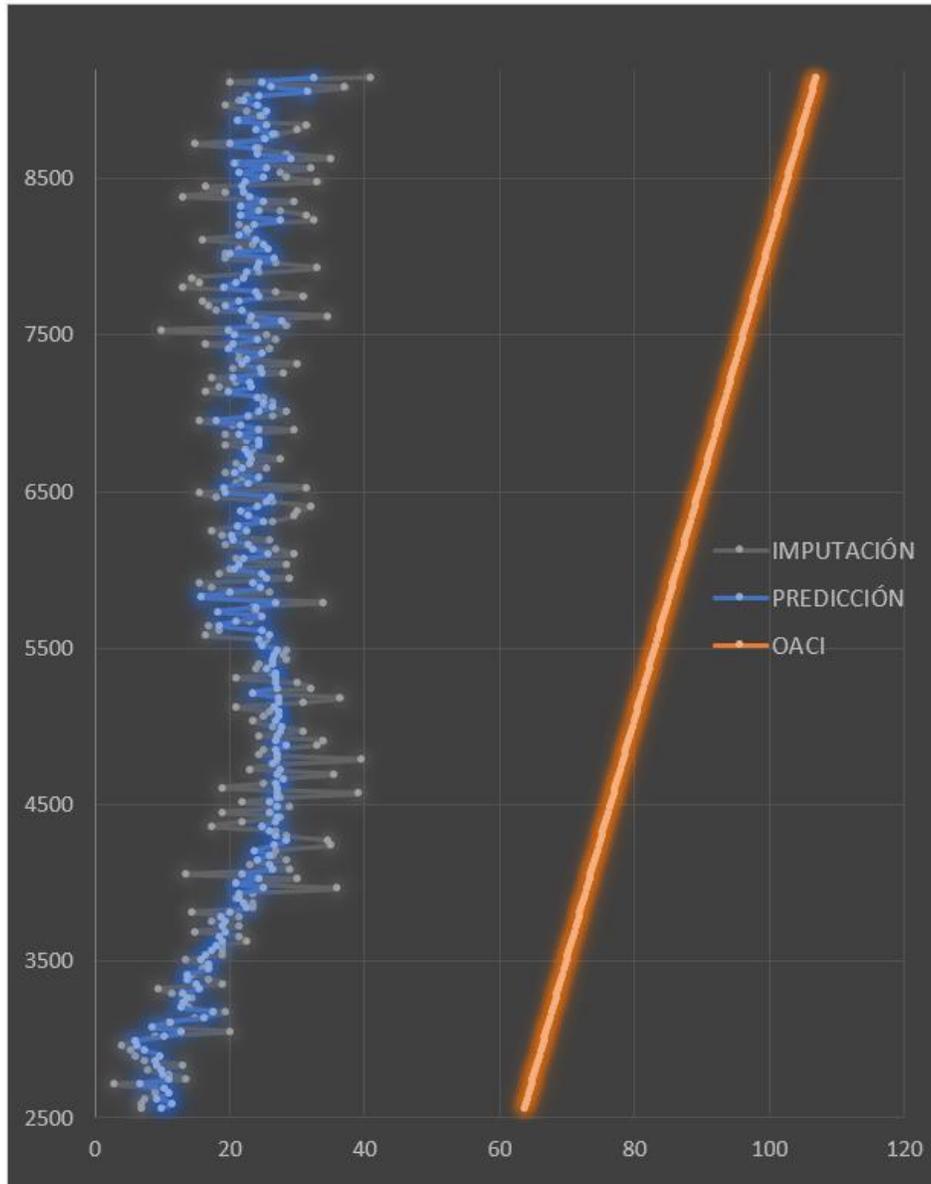
Tabla 12

Estimadores estadísticos del conjunto de datos aplicables al diseño de rutas y procedimientos de vuelo

	PRESION	ALTURA	TEMPAMB	TEMPROCIO	HUMEDAD	MEZCLAIRE	TPOTP	VIENTO
count	217.000000	217.000000	217.000000	217.000000	217.000000	217.000000	217.000000	217.000000
mean	512.417512	5852.218618	-4.101382	-8.513134	32.225806	2.375023	333.096037	34.477327
std	123.599726	1913.771765	11.006826	12.388253	22.709002	2.562588	6.447542	5.211497
min	320.750000	2560.350000	-25.300000	-35.000000	1.500000	0.030000	317.520000	18.450000
25%	406.000000	4206.270000	-12.950000	-19.050000	12.500000	0.420000	327.550000	32.580000
50%	501.850000	5852.190000	-3.500000	-7.250000	29.500000	1.260000	332.900000	35.800000
75%	615.000000	7498.200000	5.600000	2.600000	49.000000	3.130000	338.600000	38.280000
max	749.000000	9144.120000	18.900000	12.250000	86.500000	9.490000	345.030000	44.940000

Imagen 13

Predicciones a partir de imputación de variables



7. CONCLUSIONES

Los avances tecnológicos han facilitado la aceptación de nuevos métodos de captura de datos meteorológicos, particularmente el registro de datos de vuelo a bordo de aeronaves comerciales, por drones e inclusive teléfonos móviles. Esta condición significa una ventaja relacionada con la disponibilidad de información que puede ser utilizada con fines de investigación y la implementación de nuevos conceptos como el pronóstico en futuro muy cercano, conocido en inglés como *nowcasting*.

Respecto al tratamiento de datos meteorológicos, se reconocen dos metodologías importantes a nivel global. En primer lugar, los modelos de pronóstico meteorológico, los cuales hacen uso de métodos numéricos basados en ecuaciones derivadas de la mecánica de fluidos, la termodinámica y la estocástica; con el fin de generar productos a partir de procesos analíticos y que sirven para la planificación operacional de los vuelos. De otra parte, se encuentran los modelos de aprendizaje automático, los cuales, con origen en información previa sobre un fenómeno en particular, identifican patrones para realizar predicciones basadas en regresión o clasificaciones mediante relaciones con etiquetas. En general, estos modelos significan un apoyo importante en diferentes disciplinas, como las que relacionan al medio ambiente con energías renovables.

Los modelos estadísticos de predicción y clasificación han participado también en el desarrollo de iniciativas orientadas a solucionar dificultades en el transporte aéreo. El caso está documentado con relación a la identificación de patrones de viento para predecir la magnitud de sus componentes (dirección e intensidad), en apoyo a la toma de decisiones de índole operacional; por ejemplo, la decisión de seguir una ruta específica.

Con relación a la propuesta plasmada en este documento, el estudio del viento en altura asociado con el diseño de procedimientos de vuelo se presume como el primero que guarda este enfoque. Por esta razón se consolida como una referencia para futuros estudios con aplicación en otros escenarios particulares de la navegación aérea. Aquí toma importancia la geoestadística, a partir de la cual el estudio del viento con modelos estadísticos puede facilitar en mayor medida la disponibilidad de datos georreferenciados y con una resolución horizontal determinada en procura de optimizar la gestión del espacio aéreo y del tránsito aéreo.

Los datos obtenidos de la estación de radiosondeo de Bogotá corresponden en perfil a información del comportamiento de las variables meteorológicas de la troposfera (hasta 10.000 m) y la sección baja/media de la estratosfera (hasta 30.000 m). En esta última se presenta la inversión térmica que consiste en un aumento de la temperatura a mayores alturas. En un principio se evidenció que al considerar los valores de las variables correspondientes al total del perfil vertical (2.546 m a 31.000 m), la naturaleza aleatoria de las variables significó un reto en términos de precisión, rendimiento durante el proceso de entrenamiento / aprendizaje y los recursos de cómputo. Estas dificultades fueron solucionadas al momento de segmentar la base de datos para considerar únicamente la porción de la atmósfera bogotana donde evolucionan las aeronaves; es decir hasta 15.240 m o 50.000 ft. Hasta este nivel las variables presentan un comportamiento aproximadamente lineal, lo que condujo a un mejor rendimiento de los modelos diseñados.

Se utilizaron modelos estadísticos relacionados con funciones lineales y de regularización como el *RIDGE / LASSO*, árboles aleatorios como *RandomForest* y *GradientBoostingTrees*, métodos de ensamble o *stacking* para integrar dos modelos, redes neuronales y modelos de identificación de subgrupos con características comunes *K-NearestNeighbors*. Todos los modelos fueron sometidos a la misma metodología, iniciando con la estandarización de los datos para un mejor rendimiento, el desarrollo e integración de las funciones de entrenamiento y transformación, la optimización de hiperparámetros mediante procesos iterativos de entrenamiento, prueba y evaluación, y finalmente la generación de predicciones para identificar el mejor modelo a través de la métrica RMSE, donde el modelo secuencial de potenciación del gradiente, demostró la mejor performance, con un valor inicial de 8,78 nudos.

Al someter los datos al modelo, se evidenció una tendencia común en las predicciones hacia un valor aproximado de 15 nudos, que en la mayoría de los casos resultaron muy inferiores respecto al valor de las

observaciones. Esto permitió concluir que los valores obtenidos no son útiles para el diseño de rutas y procedimientos de vuelo, considerando que no representa las condiciones más críticas (valor máximo del viento). Esto, como consecuencia de la multiplicidad de valores de viento para un mismo nivel, es decir de la variabilidad de los datos y la mayor densidad representada en valores bajos e influencia de los valores de moda de los datos.

Como solución se decidió generar una base de datos nueva a partir de un único valor de altura, cada uno de los cuales resultó relacionado con valores mínimos de presión, humedad y razón de mezcla del aire, y máximos para todos los tipos de temperatura y viento. Esta información fue utilizada para evaluar nuevamente el modelo, obteniendo un RMSE de 9,24 nudos. Las predicciones demostraron que el modelo era capaz de identificar el patrón de comportamiento del viento en cuanto a cambios en su magnitud; no obstante, se presentó un error positivo máximo de 15 nudos. De manera simultánea se evaluó el rendimiento del modelo y el comportamiento de las observaciones con relación al viento OACI (originado en la ecuación $w = 2h + 47$), encontrando que, para el caso particular de la atmósfera bogotana, los valores de viento derivados de la citada ecuación son prácticamente improbables, al presentar valores que superan los registros máximos hasta en 45 nudos.

Una vez se logró identificar el comportamiento del modelo, se procedió con la imputación de valores a las variables predictoras, con el fin de obtener los insumos necesarios para el funcionamiento del modelo, logrando de esta manera la disponibilidad de datos de viento en intervalos de 30,48 m (100 ft), hasta un nivel de 9.144 m (30.000 ft). Los valores incluyen el valor de error máximo, con el fin de considerar las condiciones más críticas del viento en cada nivel, dando oportunidad a los diseñadores de procedimientos de evaluar su aplicación directa en un escenario particular u optar por el valor calculado que representa el 95% de probabilidad para cualquier nivel: 43,05 nudos.

Desde la óptica de la gestión del tránsito aéreo y particularmente la organización y gestión del espacio aéreo representada en actividades de diseño de rutas y procedimientos de vuelo, se consideran justificados los resultados que conducen a la utilización de los valores de viento obtenidos del modelo, a partir de la demostración del comportamiento del radio de viraje de las aeronaves en diferentes escenarios y su injerencia en las dimensiones de las áreas de protección. De esta manera se propone su implementación como medio para aumentar el nivel de actuación del sistema ATM, y consecuentemente como estrategia para satisfacer las expectativas de la sociedad y la comunidad internacional de aviación.

8. REFERENCIAS

- Alarcón, M. & Casas, M. C. (1999). El movimiento atmosférico. En Edicions Universidad Politècnica de Catalunya (Ed.), *Meteorología y clima* (pp. 63 – 85). Edicions UPC.
- Amat, J. (2 de septiembre de 2020). *Machine learning con Python y Scikit-learn*. https://www.cienciadedatos.net/documentos/py06_machine_learning_python_scikitlearn.html
- Aydogmus, Y., Erdal, H. & Turkan, Y. (2016). The prediction of the wind speed at different heights by machine learning methods. *Journal of optimization and control: Theories & Applications*. 6(2). DOI: 10.11121/ijocta.01.2016.00315
- Cattani, D., Eckert, P., Eggimann, F. & Kretzschmar, R. (2004). Neural Network Classifiers for Local Wind Prediction, *Journal of Applied Meteorology*, 43(5), 727-738. <https://journals.ametsoc.org/view/journals/apme/43/5/2057.1.xml>
- Chango, Ramón., Lupera, Pablo., Llugsi, R., & Suárez, A. (2017). Implementación de un Sistema Aéreo de Medición y Almacenamiento de Parámetros Meteorológicos Georreferenciados para Zonas Pequeñas. *Revista Politécnica*, 39(2), 17-26. Recuperado el 10 de agosto de 2021, de http://scielo.senescyt.gov.ec/scielo.php?script=sci_arttext&pid=S1390-01292017000200017&lng=es&tlng=es.

- Cole, R. , Benjamín, S. & Schwartz, B. (1999). Accuracy of ACARS Wind and Temperature Observations Determined by Collocation. *Journal Weather and Forecasting*, 14 (6), 1032-1038. DOI: [https://doi.org/10.1175/1520-0434\(1999\)014<1032:AOAWAT>2.0.CO;2](https://doi.org/10.1175/1520-0434(1999)014<1032:AOAWAT>2.0.CO;2)
- Crespo, E. & Zúñiga, I. (2015). *Meteorología y climatología*. Universidad Nacional de Educación a Distancia (UNED).
- Ikuchi, R., Misaka, T., Obayashi, S., Inokuchi, H., Oikawa, H. and Misumi, A. (2018), Nowcasting algorithm for wind fields using ensemble forecasting and aircraft flight data. *Meteorological Applications*, 25(3), 365-375. <https://doi.org/10.1002/met.1704>
- Kim, J., Chan, W. N., Sridhar, B., & Sharman, R. D. (2015). Combined Winds and Turbulence Prediction System for Automated Air-Traffic Management Applications. *Journal of Applied Meteorology and Climatology*, 54(4), 766-784. <https://journals.ametsoc.org/view/journals/apme/54/4/jamc-d-14-0216.1.xml>
- Organización de Aviación Civil Internacional [OACI], (2005). *Documento 9854. Concepto operacional de gestión del tránsito aéreo mundial*. Ed. 1. OACI.
- OACI, (2008). *Documento 9882. Manual sobre requisitos del sistema de gestión del tránsito aéreo*. Ed. 1. OACI.
- OACI. (2014). *Documento 9377. Manual sobre coordinación entre los servicios de tránsito aéreo, los servicios de información aeronáutica y los servicios de meteorología aeronáutica*. Ed. 6. Montreal: OACI.
- OACI. (2019). *Documento 8896. Manual de métodos meteorológicos aeronáuticos*. Ed. 12. Montreal: OACI
- OACI, (2020). *Documento 8168. Operación de aeronaves. Construcción de procedimientos de vuelo por instrumentos y visual*. Vol. II, Ed. 7. OACI.
- Ruíz, J. F., Serna, J. & Zapata, H. (2017). *Atlas de viento de Colombia*. Imprenta Nacional de Colombia.
- Salvador, R. (2015). *Experimento de Sondeo Vertical de la Atmósfera en el Valle de Cumbayá, Ecuador: Protocolo de Lanzamiento de Sondas Meteorológicas y Resultados Preliminares* [Trabajo de grado, Universidad San Francisco de Quito]. Repositorio Institucional Colegio de Ciencias e Ingeniería <http://repositorio.usfq.edu.ec/handle/23000/4144>
- Sánchez, L. A. (2016). *Prototipo de radiosonda meteorológica*. [Tesis de grado, Instituto Politécnico Nacional]. Repositorio institucional <https://tesis.ipn.mx/bitstream/handle/123456789/21275/%E2%80%9CPrototipo%20de%20Radiosonda%20Meteorol%C3%B3gica.pdf?sequence=3&isAllowed=y>
- Waller, J. A. (2020). The use of unconventional observations in numerical weather prediction. *Meteorological applications Journal*. <https://rmets.onlinelibrary.wiley.com/doi/epdf/10.1002/met.1948>