



**Análisis de Causas de la aparición de
Inclusiones en la Fabricación de Aceros a
partir de Variables de Producción.**

**Analysis of Causes of Inclusion Formation in Steel
Manufacturing from Production Variables.**

Trabajo de Fin de Máster

Máster en Data Science

Universidad de Cantabria

Director: Diego Ferreño Blanco

Codirector: Fernando Aguilar Gómez

Autora: Carla King Molines

Fecha: 21 de junio de 2024

Índice

1	Introducción.	4
2	Proceso de producción del acero.	5
3	Origen y detección de inclusiones.	8
4	Datos y pretratamiento.	9
5	Modelado	11
5.1	Modelos empleados.	11
5.2	Evaluación de los modelos.	14
6	Resultados.	16
6.1	Análisis exploratorio.	16
6.2	Modelización.	18
6.2.1	Modelización con variable objetivo binaria.	18
6.2.2	Modelización con variable objetivo con tres categorías.	22
7	Conclusiones.	29

Resumen

Las inclusiones no metálicas pueden generar un impacto negativo en las características del acero final, estos efectos perniciosos son particularmente relevantes en el caso de los aceros especiales. En este tipo de aceros, tratar de reducir, tanto el número, como el tamaño de las inclusiones se vuelve un aspecto crítico. Para ello, el primer paso es comprender las variables que desempeñan un papel relevante en su formación. Esto, sin embargo, supone un gran desafío, dado que el número de variables implicadas en el proceso productivo es muy elevado.

En este trabajo se van a emplear técnicas de *data mining* y aprendizaje automático para tratar de lidiar con esta alta dimensionalidad. Se han ajustado diferentes algoritmos, empleando Optimización Bayesiana para el ajuste de los hiperparámetros, hasta encontrar los modelos que presentan un mejor desempeño. A continuación, se han extraído las características que desempeñan un papel más relevante en la predicción del tamaño de inclusión.

El análisis se ha realizado tanto con una variable de clasificación binaria como ternaria. Tras ajustar los modelos, los mejores resultados se obtienen con XGBoost para el caso de la clasificación en tres clases y con Gradient Boosting en la clasificación binaria. Las variables que permiten clasificar mejor las instancias son *Temperatura Artesa MÁXIMA ($^{\circ}C$)* y *Velocidad MÍNIMA (m/min)*

Los resultados demuestran que, incluso en casos donde se dispone de un número reducido de instancias, el empleo de técnicas de machine learning permite extraer información relevante de los datos.

Palabras clave: Inclusiones no metálicas, Acero, Aprendizaje automático, Data mining, Análisis de datos

Abstract

Non-metallic inclusions can have a negative impact on the characteristics of the final steel product, and these detrimental effects are particularly relevant in the case of special steels. In this type of steel, reducing both the number and size of inclusions becomes a critical aspect. To achieve this, the first step is to understand the variables that play a relevant role in their formation. However, this poses a great challenge, given that the number of variables involved in the production process is very high.

In this work, data mining and machine learning techniques will be employed to deal with this high dimensionality. Different algorithms have been adjusted, using Bayesian optimization, until finding the models that present the best performance. Subsequently, the features that play a more relevant role in predicting the size of inclusions have been extracted.

The analysis has been conducted using both a binary and a ternary classification variable. After adjusting the models, the best results are obtained with XGBoost for the case of classification into three classes and with Gradient Boosting in binary classification. The variables that best classify the instances are "Maximum Artisan Temperature (°C)" and "Minimum Speed (m/min)".

The results demonstrate that, even in cases where a reduced number of instances are available, the use of machine learning techniques allows relevant information to be extracted from the data.

Keywords: Non-metallic inclusions, Steel, Machine learning, Data mining, Data analysis

1 Introducción.

La producción de acero en el año 2023 alcanzó las 1.892 millones de toneladas y el valor añadido aportado, directa o indirectamente, por este sector representa el 3.8% del PIB mundial [1]. Esto convierte a la industria del acero en una de las más relevantes a nivel global, junto al petróleo y el gas.

El acero es un material imprescindible para múltiples sectores, desde la construcción de infraestructuras hasta la fabricación de bienes de consumo. Uno de los aspectos críticos en su fabricación es el asegurar la calidad en el producto final, para ello, es imprescindible reducir en la medida de lo posible el número y tamaño de las inclusiones.

Las inclusiones son compuestos no metálicos que se encuentran en el interior del acero y pueden afectar negativamente a su resistencia a la tracción y a la fatiga[2], a su tenacidad y ductilidad[3] y a su resistencia a la corrosión[4] [5]. Es por ello por lo que, reducir su aparición y su tamaño es crítico en la fabricación de aceros de alta calidad.

Determinar las causas de estas inclusiones supone un gran desafío dada la complejidad del proceso productivo y el gran número de variables que en él se ven implicadas. Sin embargo, la introducción de técnicas de *machine learning* posibilita el análisis de estas complejas interacciones entre variables.

En el presente trabajo se han aprovechado las posibilidades que proporcionan los algoritmos de *data mining* y *machine learning* para comprender el papel que desempeñan las diferentes variables de producción en el tamaño de las inclusiones.

Sin embargo, alta dimensionalidad del dataset del que se dispone supone un desafío en términos la generación de modelos capaces de generalizar. Puesto que se dispone de un número muy reducido de observaciones y un gran número de variables, el uso de las técnicas usualmente empleadas en este tipo de problemas llevaría a un sobreajuste de los modelos, esto es, que los modelos aprenderían directamente los datos y no los patrones inherentes a ellos. Es por ello por lo que el proceso de selección y ajuste de los modelos resulta en este caso de especial relevancia.

El trabajo se ha dividido en seis apartados, en el primero de ellos se describe el proceso de producción del acero. Seguidamente se abordan los orígenes de las inclusiones y la

forma de detectarlas. En el apartado 4 se explica la base de datos y el pre-tratamiento que se ha aplicado en ella. En el quinto apartado se exponen los modelos y la metodología de evaluación que se ha empleado. Finalmente se exponen los resultados, en primer lugar utilizando una variable objetivo binaria y en segundo lugar para una clasificación en tres clases.

2 Proceso de producción del acero.

En la actualidad existen principalmente dos tipos de acerías, las acerías de horno de oxígeno básico (BOP) y las acerías de horno de arco eléctrico (EAF).

El proceso realizado en las BOP se denomina siderurgia integral ya que la fabricación del acero parte del arrabio proveniente de la fundición del mineral de hierro en los altos hornos. El arrabio está constituido por hierro líquido, carbono, silicio, manganeso y azufre.

El objetivo principal del horno de oxígeno básico es reducir el contenido de carbono y eliminar las impurezas que contiene el arrabio. Esto se hace principalmente mediante la inyección de oxígeno, la cual provoca la oxidación del carbono, liberando monóxido y dióxido de carbono [6]. Asimismo, el oxígeno inyectado reacciona con las impurezas, dando lugar a la formación de óxidos, los cuales flotan a la escoria. Dado que este conjunto de reacciones es exotérmico no es necesario aplicar otra fuente de energía para mantener el proceso.

Todo este proceso resulta en un acero con bajo contenido de carbono, el cual será refinado, en la fase de metalurgia secundaria, hasta alcanzar la composición deseada.

Por otro lado, en las acerías de arco eléctrico, las materias primas que se emplean son principalmente chatarra y mineral prerreducido[7]. Este tipo de acerías consta de tres instalaciones principales: el horno eléctrico, el horno cuchara y las líneas de colada continua[8].

En primer lugar, la materia prima se selecciona, en función de la calidad que se pretende alcanzar, y se carga en el horno eléctrico. El horno de arco eléctrico recibe este nombre debido a que el proceso de calentamiento se realiza a través de un conjunto de electrodos,

típicamente tres, entre los cuales se genera un arco eléctrico que calienta su contenido. Por tanto, el calentamiento en el caso de los hornos EAF sí que requiere de energía eléctrica, a diferencia que en los BOP.

El horno está compuesto de tres partes, la solera, la bóveda y la cuba. Las tres contienen un revestimiento de material refractario para proteger la estructura del horno de las altas temperaturas.

La solera está en la parte inferior del horno y es donde se alojará el acero líquido. Además del revestimiento de material refractario, contiene un revestimiento de desgaste que es el que está en contacto con el acero.

La cuba se sitúa debajo de la solera y está constituida por paneles refrigerados, además contiene aberturas por donde se pueden tomar mediciones.

La bóveda se sitúa en la parte superior del horno y contiene los orificios por donde se insertan los electrodos de grafito. Consta de otros dos orificios, uno para permitir la eliminación de los gases y otro para añadir los escorificantes.

El proceso en el EAF incluye también la adición de oxígeno, esto ayuda a eliminar las impurezas, a reducir el tiempo de fusión y a formar la escoria, la cual ayudará a prevenir la disipación de calor, así como a reducir el consumo de refractario.

A continuación se presenta una representación esquemática de un horno EAF.

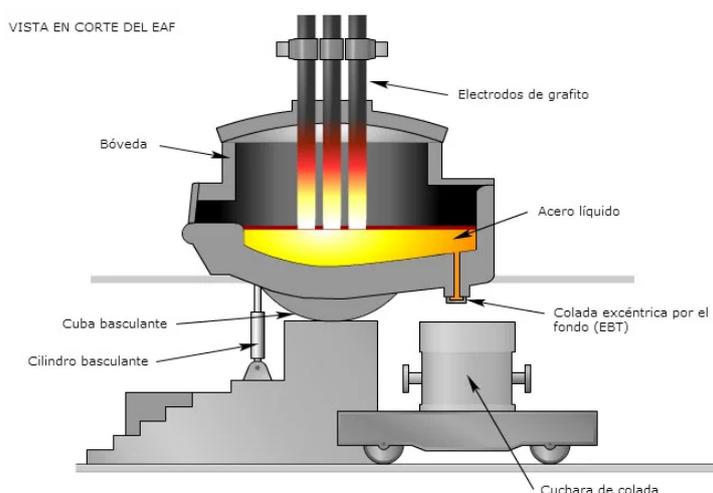


Figura 1: Esquema de un horno de arco eléctrico (Fuente: <https://metfusion.wordpress.com>)

Una vez realizada la fundición en el EAF la colada pasará al horno cuchara donde se transportará el acero fundido desde el EAF a la colada continua. Este horno constituye la parte más importante del proceso en cuanto a la limpieza incursionaría. En su interior, se mantienen condiciones de vacío, lo que reduce la formación de inclusiones relacionadas con la exposición al aire del material fundido. Durante el periodo en el que el material se aloja en el horno cuchara se realiza la desoxidación del acero, el ajuste de la composición de sulfuros, el control de la temperatura, la limpieza y la adición de elementos para conseguir las características deseadas en el acero final.

En el horno cuchara el proceso de calentamiento también se realiza mediante unos electrodos de grafito, en este caso de menor tamaño. Para evitar el sobrecalentamiento y favorecer que la temperatura sea uniforme se debe agitar el contenido de la cuchara, en el caso de GSW el método que se emplea es el de inyectar argón desde el fondo de la cuchara.

Finalmente, cuando el proceso de tratamiento en la cuchara finaliza, la colada pasará a la artesa, que es el recipiente que contiene el acero antes de que este pase a las líneas de colado. El traspaso del contenido de la artesa a los moldes se realiza a través de la buza. Existen distintos tipos de buza que se escogerán en función de los requerimientos de calidad, cuando la exposición a oxígeno no supone un problema se puede emplear la buza calibrada. Sin embargo, en el caso de aceros con alto contenido en aluminio se requiere del uso de buza sumergida, la cual evita que se produzca la reoxigenación y con ella la generación de inclusiones de alúmina.

En esta etapa es especialmente relevante la temperatura de la artesa, así como la velocidad a la que entra el acero en los moldes.

Una vez depositado el acero en los moldes este pasará a zona de refrigeración secundaria donde se controlará el proceso de enfriamiento hasta la solidificación completa.

3 Origen y detección de inclusiones.

Las inclusiones en los aceros pueden dividirse en dos grupos, endógenas o exógenas. Las endógenas están compuestas por elementos como los óxidos, sulfuros, nitruros o fosfuros y su origen está en las propias reacciones del proceso de fabricación. Por otro lado, las exógenas provienen normalmente del desgaste de alguno de los materiales, como los refractarios, o de la propia escoria[9].

Las inclusiones endógenas son más abundantes, pero, debido a su reducido tamaño, el efecto que tienen en las propiedades finales del acero es menos nocivo. En cambio, las inclusiones exógenas, aunque más escasas, tienen un efecto más dañino sobre el producto final, puesto que pueden reducir la resistencia a la fatiga del material de modo que este se vuelve más susceptible a fracturas [10].

Debido a estos efectos perniciosos sobre el acero, es fundamental reducir, en la medida de lo posible, el tamaño de las inclusiones en el producto final.

En la actualidad existen diversas metodologías para caracterizar el contenido inclusionario. Mediante microscopía óptica o electrónica de barrido, mediante ultrasonidos, empleando partículas magnéticas, entre otras. Sin embargo, estas metodologías no son óptimas para la detección macroinclusiones causantes de fallo por fatiga.

En este caso se emplea una metodología novedosa desarrollada en el LADICIM que permite detectar este tipo de inclusiones de una forma más eficiente y efectiva que las empleadas normalmente en la industria.

Los datos promedio del tamaño de inclusión han sido obtenidos empleando esta metodología y cada una de las 49 instancias proviene de la realización de 30 ensayos.

El método desarrollado por LADICIM presenta varias ventajas frente a los ensayos por fatiga típicamente empleados en la industria, la principal es que el proceso es mucho más rápido, además permite detectar inclusiones causantes de rotura, no solo en la parte exterior de la probeta, como ocurre en el caso de los ensayos por flexión rotatoria, sino también en la región central[12].

4 Datos y pretratamiento.

La base de datos a partir de la que se ha realizado el análisis consta de 108 variables obtenidas a partir de mediciones tomadas por la empresa en las distintas fases de la fabricación de los aceros. A partir de estas se va a obtener la información para predecir las variables objetivo.

Las variables objetivo se han obtenido experimentalmente a partir de muestras de acero proporcionadas por la empresa. Para cada una de ellas se ha realizado un análisis del contenido inclusionario siguiendo la metodología descrita en el apartado 2. Con ello, se ha obtenido información relativa al número de inclusiones, su tamaño y composición.

De modo que, para cada una de las muestras de acero pertenecientes a las distintas coladas, se han realizado 30 ensayos. A partir de los resultados obtenidos en estos, se ha promediado el tamaño de inclusión para cada muestra, resultando en un total de 49 instancias.

Por tanto, la base de datos inicial consta de 108 variables predictoras y de 49 observaciones. Esto resulta en un bajo ratio observación-variable, lo que supone grandes desafíos en cuanto a la obtención resultados fiables. Esta característica aumenta el riesgo de sobreajuste y dificulta la identificación de patrones.

Para paliar estos efectos se ha optado por emplear validación Leave-One-Out. Este tipo de validación cruzada consiste en emplear todas las observaciones, excepto una, como conjunto de entrenamiento, utilizando la restante para realizar la validación. Este proceso se repite para todo el conjunto de observaciones, calculando el error para cada división train-test. Finalmente, el error del modelo se calcula a partir del promedio de los errores de cada una de las divisiones.

Este tipo de validación supone un alto coste computacional, pero permite reducir la variabilidad derivada del sobreajuste del modelo a los datos de entrenamiento, la cual, en el contexto de bases de datos con pocas observaciones y muchas variables, como es este caso, es especialmente relevante.

Se ha modelizado el problema a través de una clasificación binaria (separando por la mediana) y de una clasificación ternaria (separando en terciles). En el caso de la división

simple la variable objetivo tomará valor 0 cuando el tamaño de inclusión promedio se sitúa por debajo de la mediana y 1 en caso contrario. En el caso de la división ternaria, la variable objetivo se obtiene partir de la división de las muestras en terciles en base a su tamaño de inclusión promedio. Por consiguiente, tomará valor 0 en el caso de que el tamaño de inclusión se encuentre en el primer tercil, 1 en el caso de encontrarse en el segundo tercil y 2 en el caso de pertenecer al tercer tercil.

Durante el proceso de ingeniería de características se ha barajado los datos de modo que los modelos no aprendan a clasificar las observaciones en base a la posición que ocupan en el dataset.

Las variables de tipo temporal se han convertido a minutos. Las variables categóricas con únicamente dos valores se han transformado en variables binarias. Para aquellas con más de dos valores, se ha aplicado *one-hot-encoding* para crear variables dummies para cada categoría.

Asimismo, se han eliminado aquellas variables que no aportaban información por ser constantes para todas las muestras, así como aquellas que presentaban un gran número de valores faltantes.

Para las variables que presentaban un número de valores faltantes minoritario con respecto al total de observaciones, se ha imputado el valor de la variable utilizando la media.

Finalmente, para asegurar que los resultados no se vean afectados por la magnitud de las variables se ha escalado el conjunto de predictores, esto es, estandarizar las características de modo que tengan media nula y varianza unitaria.

5 Modelado

5.1 Modelos empleados.

La modelización se ha realizado empleando Python, con las librerías numpy, pandas, scikit-learn, shap, scikit-optimize, xgboost, scipy, matplotlib y seaborn.

Para valorar la consistencia de los resultados se han empleado diversos modelos de Machine Learning.

- **Regresión logística multinomial.** Se ha empleado como punto de partida para poder comparar con el resto de modelos. Este tipo de regresión utiliza la función sigmoide para predecir la probabilidad de pertenencia a una determinada categoría y ajusta el modelo en base a la función de pérdida de entropía cruzada, la cual mide la diferencia entre las probabilidades predichas y las clases observadas.

Sin embargo, para el tipo de problema que se pretende abordar, con alta dimensionalidad, se espera que un modelo basado en la regresión logística no ofrezca los mejores resultados. Este tipo de modelos se ve afectado negativamente por la presencia de multicolinealidad en los datos, la cual, en el caso de tener un gran número de variables predictoras, puede ser un problema.

- **K-Nearest Neighbors (KNN).** Este algoritmo se basa en clasificar las instancias en base a las distancias entre las características, de modo que se asignará una nueva observación a una determinada clase en base al “voto por mayoría” de los k vecinos más cercanos.

Este algoritmo puede presentar problemas para calcular la distancia entre las instancias cuando el conjunto de datos presenta alta dimensionalidad. Esto es conocido como la “Maldición de la dimensionalidad”, la cual dificulta el identificar los vecinos más cercanos. Por este motivo, es esperable que este algoritmo pueda presentar un desempeño pobre para el dataset de estudio.

- **Clasificador Bayesiano Ingenuo (Naive-Bayes).** Se trata de un algoritmo de clasificación probabilística basado en el Teorema de Bayes. Este algoritmo asume el supuesto de independencia entre variables y calcula las probabilidades en base a la probabilidad condicional de cada característica dada una determinada clase, por

tanto, cada variable contribuye a la predicción de forma independiente.

A pesar de que en este caso el supuesto de independencia puede no resultar válido, es interesante valorar el desempeño de este algoritmo dada la ventaja que presenta en cuanto a la reducida cantidad de datos que requiere para la estimación de los parámetros.

- **Arboles de clasificación.** Este tipo de modelo busca clasificar las observaciones mediante la sucesiva división del conjunto de datos, estas divisiones se realizan en base a las características o umbrales que permiten discriminar entre clases, de modo que se maximice la homogeneidad de cada uno de los subconjuntos.

El uso de modelos basados en árboles de decisión es especialmente interesante en este caso ya que el propio modelo realiza una selección de características, dejando fuera aquellas que son menos relevantes para la clasificación. Además, este tipo de modelo tiene la ventaja de tener una sencilla interpretación.

- **Random Forrest.** Se trata de una técnica de *bagging*, esto es, que utiliza la unión de varios modelos débiles entrenados en paralelo. En este caso usa concretamente un conjunto de árboles de clasificación, gracias a ello se consigue mitigar el principal problema de los árboles sencillos, que es la sensibilidad que presentan al ruido.
- **Gradient Boosting.** Del mismo modo que en el caso anterior, este algoritmo entrena un conjunto de modelos débiles, pero en este caso, al tratarse de una técnica de boosting, no se realiza en paralelo sino de forma secuencial, de modo que cada nuevo modelo intenta mejorar el anterior. En el caso concreto de Gradient Boosting, esto se hace mediante reajustar el nuevo modelo de modo que minimice la función de pérdida.
- **Extreme Gradient Boosting (XGBoost).** Este algoritmo se basa en los principios de Gradient Boosting pero XGBoosting está optimizada para asegurar una mejor eficiencia y rendimiento.
- **Adaptative Boosting (AdaBoost).** Se trata de otro método de boosting, a diferencia que en el caso de *Gradient boosting*, *Adaboost* mejora progresivamente los modelos mediante el ajuste de los subsiguientes modelos a las instancias mal clasificadas.

Los modelos de boosting, tanto *Gradient Boosting* como *Adaboost*, se centran en reducir el sesgo, es por ello que pueden generar un mayor overfitting que modelos de bagging, como Random Forest, que se centran en reducir la varianza.

- **Support Vector Machines (SVM).** Este tipo de algoritmo busca separar las clases mediante un hiperplano óptimo. Se define como hiperplano óptimo aquel que maximiza el margen, esto es, la distancia entre el hiperplano y los puntos más cercanos de las diferentes clases.

La principal ventaja de este método es que permite trabajar con espacios de alta dimensionalidad y puede funcionar bien incluso con un número limitado de observaciones.

- **Red neuronal.** Consta de tres tipos de capa, la de entrada, las ocultas y la de salida. En cada una de estas capas habrá un determinado número de nodos a los que se denomina neuronas. Estas neuronas están conectadas a otras, formando una red. Cada una de estas conexiones tiene asignado un peso, el cual puede ser considerado como la ponderación de la importancia que asigna la red a cada instancia, y una función de activación, la cual solo permitirá el paso de información si se supera cierto umbral. De este modo, la salida de cada capa, obtenida a partir de aplicar la función de activación a los pesos, pasará a la siguiente, repitiéndose el proceso en todas las sucesivas capas.

A pesar de que las redes neuronales son una buena opción para capturar patrones complejos, pueden no ser la mejor opción en este caso, puesto que suelen requerir de una gran cantidad de datos para ser entrenadas y tienden al sobreajuste.

5.2 Evaluación de los modelos.

Para evaluar y comparar los modelos se ha empleado principalmente la métrica de accuracy. Esta mide la proporción de aciertos en la predicción del modelo, es decir, del total de casos predichos, cuantos se corresponden con el verdadero valor. La ecuación que describe la accuracy se presenta a continuación.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \quad (1)$$

Donde TP, TN, FP, FN representan los verdaderos positivos, verdaderos negativos, falsos positivos y falsos negativos, respectivamente.

Dado que se va a realizar una clasificación tanto binaria como ternaria, cabe destacar que la valoración del desempeño no es igual en ambos casos. En el caso de la clasificación con variable objetivo binaria, la accuracy que cabría esperar de una clasificación aleatoria es de en torno a 0,5. En cambio, en el caso de la división en tres clases, la accuracy esperada de una clasificación aleatoria sería en torno a 0.33. Por tanto, estos serán los valores de referencia para valorar el desempeño de los modelos en cada caso.

Otras métricas relacionadas son *precision* y *recall*. La primera describe el total de verdaderos positivos sobre el total de instancias positivas predichas. El recall, en cambio, representa el total de verdaderos positivos sobre el total de positivos predichos. Vienen descritas por las ecuaciones que se presentan a continuación.

$$Precision = \frac{TP}{TP + FP} \quad (2)$$

$$Recall = \frac{TP}{TP + FN} \quad (3)$$

Además, para comprender en mayor medida la distribución de las predicciones se ha obtenido la matriz de confusión para cada modelo. Esta representa los valores reales junto a los valores predichos, de modo que, los valores situados en la diagonal de la matriz de confusión serán los casos en los que el modelo ha acertado en la clasificación.

Una de las herramientas de evaluación grafica que se ha empleado es la curva ROC (*Receiver Operating Characteristic*) y su métrica asociada AUC (*Area Under the Curve*).

La representación de la curva ROC permite apreciar, de una forma visual, la relación entre la tasa de verdaderos positivos y la tasa de falsos positivos. El desempeño de un modelo de clasificación será mayor cuanto más arriba y a la izquierda se sitúe la curva, esto implica una alta tasa de verdaderos positivos y una baja tasa de falsos positivos.

La métrica normalmente asociada a la curva ROC, la AUC, es una representación numérica de cuan arriba a la izquierda se sitúa la curva. En una clasificación perfecta el valor que toma la AUC es de 1, en cambio, si la curva se sitúa sobre la diagonal del gráfico, el área debajo de la curva será 0.5, o lo que es lo mismo, la clasificación será equivalente a la que se produciría por azar.

Por otro lado, para comprender cómo varía la predicción en función del valor que toman las características, se han representado gráficos de dependencia parcial (PDP). Estas muestran el efecto marginal que tiene una variable determinada en la predicción final del modelo [13].

Los valores de todas las métricas anteriormente mencionadas se han obtenido empleando validación cruzada *Leave-One-Out*. En cuanto a la selección de los hiperparámetros de los modelos, se han empleado dos metodologías para comparar su desempeño. En primer lugar utilizando una búsqueda en cuadrícula (*grid-search*), esta consiste en probar diferentes configuraciones de parámetros hasta encontrar la que proporciona los mejores valores para una métrica de referencia, en este caso, la accuracy. En segundo lugar, optimización bayesiana, esta incorpora información de las iteraciones anteriores para ajustar las siguientes búsquedas, de modo que el proceso de ajuste de los hiperparámetros es más eficiente.

Finalmente, dado que el objetivo es comprender que variables desempeñan un papel relevante en el tamaño de inclusión e intentar reducir el número de variables, se ha obtenido la importancia de las características siguiendo diferentes metodologías, la importancia de características basada en impureza, la importancia en base a permutación y en base a los valores aditivos de Shapley.

Feature importance está específicamente diseñado para modelos basados en árboles y se trata de un método fundamentado en el concepto de impureza. En el contexto de los árboles de decisión la impureza es una medida de cuan “puros” son los nodos resultantes

de una determinada división, esto es, aquellos que presenten una menor mezcla de clases en sus nodos hijos. Por tanto, este algoritmo, definirá como más relevantes aquellas características que permitan separar mejor entre clases.

Permutation importance mide, una vez entrenado el modelo, como varia el rendimiento cuando se permuta aleatoriamente una de las características, de modo que las variables que al ser permutadas generan una mayor disminución en el rendimiento son consideradas como las más importantes.

Por último, SHAP (SHapley Additive exPlanations) es un método basado en los valores Shapley, los cuales tienen su fundamento en la teoría de juegos y consisten en obtener las contribuciones marginales de cada característica al resultado final, una vez consideradas todas las posibles permutaciones de características [14]. La diferencia es que SHAP computa el conjunto de los valores Shapley para el conjunto de predicciones para obtener una estimación de la contribución de cada variable a el output global del modelo.

6 Resultados.

6.1 Análisis exploratorio.

Dado que se dispone de un número limitado de observaciones es crucial intentar reducir el número de variables antes de intentar ajustar los modelos. Para ello, en primer lugar, se ha representado el conjunto de variables junto al tamaño medio de inclusión, de este modo se puede valorar si existen variables que no guarden relación con la variable objetivo. En el eje de ordenadas se ha representado la variable objetivo, tamaño medio de inclusión, medido en micras, y en el eje de abscisas se han representado los valores que toma la variable para cada caso.

En los gráficos que se presentan a continuación se representan algunas de las variables que parecen presentar variaciones significativas con respecto al tamaño de inclusión como son el vuelco, el tiempo total, la temperatura máxima de la artesa o la velocidad mínima. Sin embargo, existen un gran número de variables que parecen tener cierta relación con

la variable objetivo por lo que no es posible extraer conclusiones definitivas a partir de estos gráficos.

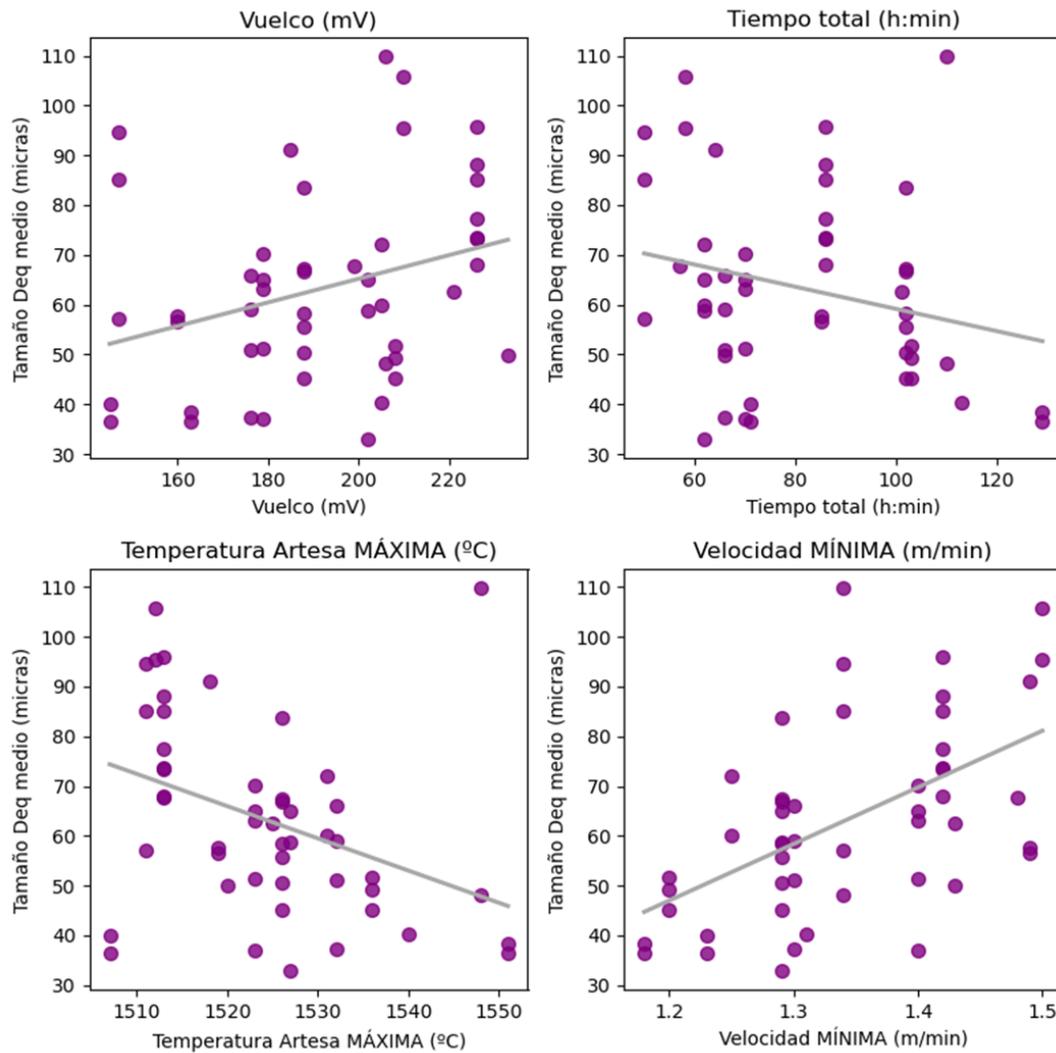


Figura 2: Gráficos de regresión de las variables en relación al tamaño de inclusión medio.

Mediante este análisis exploratorio inicial se han podido detectar un gran número de variables que parecen no fluctuar con el tamaño de inclusión, intuitivamente se puede deducir que estas no desempeñarán un papel relevante en la clasificación posterior.

Aquellas que claramente no presentan variaciones significativas o bien no aportan información, han sido eliminadas. Esto resulta en un total de 36 variables.

6.2 Modelización.

6.2.1 Modelización con variable objetivo binaria.

A partir de esta primera selección de variables se ha modelizado empleando los algoritmos descritos en el apartado 4. En este primer apartado se ha convertido la variable objetivo en una variable binaria, la cual toma valor 0 en el caso de que el tamaño de inclusión medio se sitúe en la mitad inferior de la distribución o 1 en caso contrario.

Los mejores resultados se obtienen con un árbol de decisión o con Random Forest. Estos modelos alcanzan una accuracy de 0.71. Al analizar las variables más importantes se observa que reparten la importancia entre un gran número de características. Estos resultados parecen apuntar a que se está capturando el ruido en los datos, lo que puede generar un efecto negativo en la capacidad de generalización de los modelos.

Para intentar reducir este efecto se ha realizado una selección de variables para los modelos especificados en el apartado anterior empleando *SequentialFeatureSelector* de sklearn. En los tres casos, tras la primera selección de variables, los modelos mejoran su rendimiento, esto confirma la hipótesis de que los modelos iniciales se estaban viendo afectados negativamente por el ruido asociado con la inclusión de variables irrelevantes. Además, existen un conjunto de variables que se repiten en la selección de los tres modelos.

Del mismo modo que en el caso anterior, se han eliminado las variables no seleccionadas en ninguno de los modelos y se ha repetido iterativamente este proceso hasta que la reducción de variables ya no supusiera una mejora en la accuracy.

Durante este proceso las variables seleccionadas van convergiendo hasta coincidir entre modelos. El punto en el que se alcanza una mayor accuracy es con únicamente dos variables, temperatura máxima de la artesa y velocidad mínima.

Esto no implica que el resto de las variables no desempeñen un papel relevante, pero, dado el reducido número de observaciones de las que se dispone, parece que, al introducir más variables en los modelos, estos se ven gravemente afectados por el ruido.

Los resultados obtenidos se resumen en la siguiente tabla.

Modelo	Accuracy
Regresión Logística	0.73
KNN	0.73
Árbol de Clasificación	0.73
Random Forest	0.79
Gradient Boosting	0.81
XGboost	0.77
AdaBoost	0.69
SVM	0.73
Naive Bayes	0.73

Cuadro 1: Accuracy para diferentes modelos de clasificación.

El modelo que ofrece los mejores resultados es Gradient Boosting, con una accuracy de 0.81. A continuación se presenta la curva ROC para los resultados de este modelo.

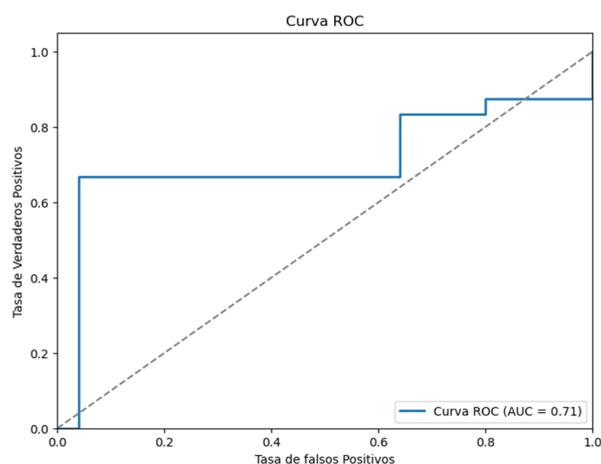


Figura 3: Curva ROC para clasificación binaria

Como se puede apreciar la curva se sitúa por encima de la diagonal hasta alcanzar una TPR de más de 0.8, tras esto pasa a situarse por debajo de la diagonal, es decir, a partir de cierto umbral el modelo experimenta una disminución en el rendimiento. Para este modelo, el AUC toma un valor de 0.71, esto representa un rendimiento medio.

Este modelo clasifica correctamente 40 de 49 instancias como se aprecia en la matriz de confusión que se presenta a continuación.

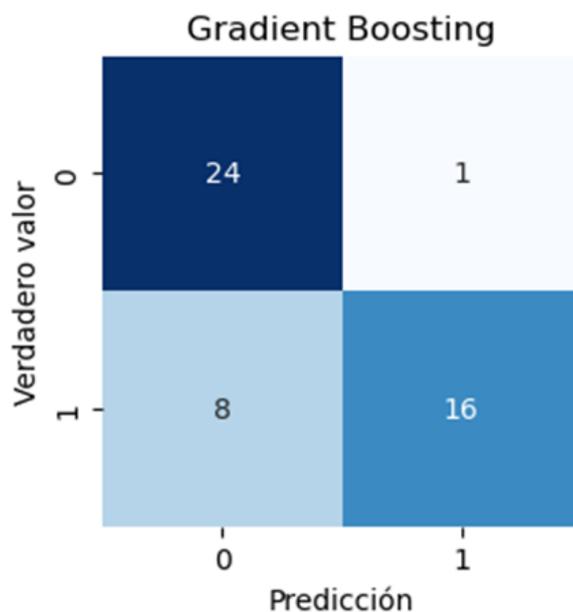


Figura 4: Matriz de confusión para clasificación binaria.

De esta matriz de confusión se pueden extraer los valores de *precision* y *recall*. La *precision* es la tasa de verdaderos positivos sobre el total de predicciones positivas y en este caso es de 0.94. El *recall* representa la tasa de verdaderos positivos sobre el total de casos positivos reales y es de un 0.66. Este valor indica que en este modelo existe un mayor número de falsos negativos que de verdaderos positivos. En el grafico que se presenta a continuación se han representado las curvas de *precision-recall* junto con el *threshold*. Esto puede ayudar a identificar si existe un *threshold* alternativo que maximice estas métricas.

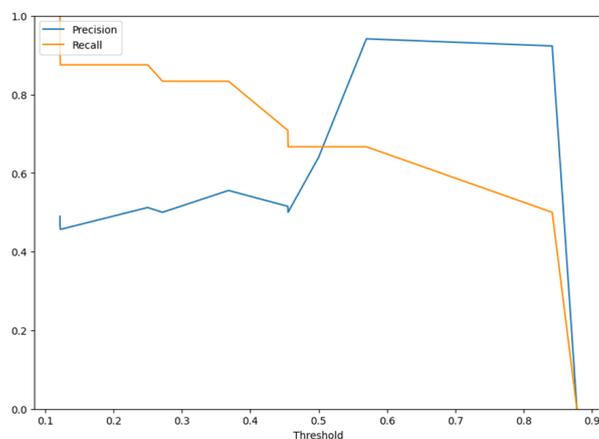


Figura 5: Curva Precision-Recall para clasificación binaria.

En este caso concreto el punto donde se cruzan *precision* y *recall* es justo en 0.5. Esto probablemente se debe a como se han estimado los hiperparametros del modelo. Ya que, por defecto, los modelos basados en árboles establecen un *threshold* de 0.5 y la optimización se ha realizado para mejorar la clasificación bajo ese umbral predeterminado. El hecho de que se situó justo en 0.5 es una muestra del buen ajuste de los hiperparametros realizado por la optimización bayesiana.

Por otro lado, si se representan las Partial Dependence Plots para las dos características, las gráficas presentan dos saltos en ambas variables. A partir de un determinado valor se produce una variación brusca, seguida de una estabilización intermedia a la que sigue otra variación. Esto puede estar reflejando una estructura subyacente en los datos que no está siendo capturada por la división binaria. Parece existir una clase adicional la cual podría representar las inclusiones de tamaño intermedio.

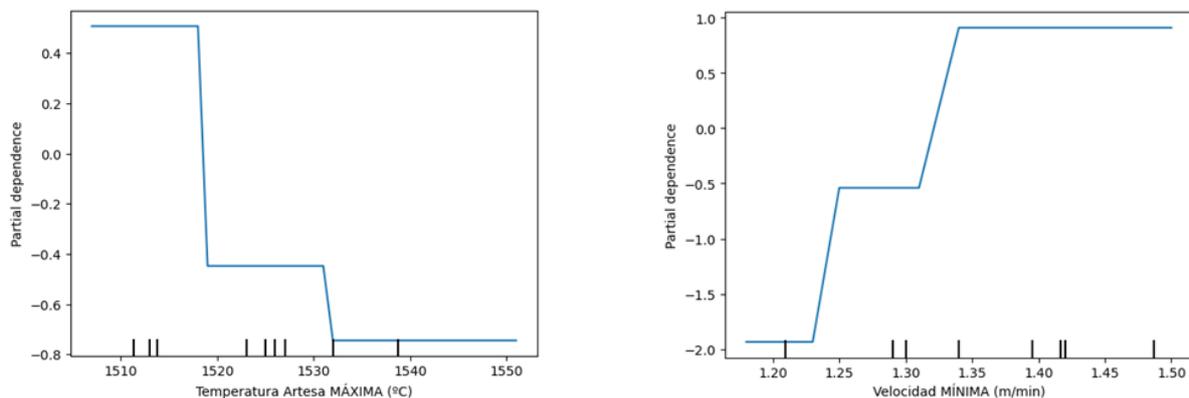


Figura 6: Partial dependence plots para clasificación binaria

En el eje *y* se representa la probabilidad de que la variable objetivo tome valor 1, es decir, que pertenezca a la clase de mayor tamaño de inclusión. Por tanto, según estos gráficos, cuando la temperatura máxima de la artesa es menor la predicción del modelo será que la inclusión sea de mayor tamaño. Por otro lado, el comportamiento contrario se aprecia en el caso de la velocidad mínima, la cual presenta unas probabilidades mayores de que la inclusión pertenezca a la clase 1, esto es, sea de mayor tamaño, a medida que la velocidad mínima aumenta.

Los puntos a partir de los cuales la predicción del modelo cambia son 1518 °C, para el primer salto, y 1532 °C para el segundo. En el caso de la velocidad los cambios se

producen en 1.23 m/min y en 1.31 m/min.

6.2.2 Modelización con variable objetivo con tres categorías.

Para este segundo apartado se ha dividido la variable objetivo en tres clases, en base a los terciles de la distribución del tamaño medio de inclusión. Esta división tiene como objetivo explorar en mayor medida el comportamiento de las PDP observado en el apartado anterior y concretar con mayor precisión los valores a partir de los cuales las variables seleccionadas tienen un efecto notable en el tamaño de inclusión.

Para comprobar que los resultados obtenidos en el apartado anterior son consistentes se ha realizado el mismo procedimiento iterativo de eliminación de variables.

En primer lugar, se ha estimado el conjunto de modelos con el conjunto de 36 variables fruto de la primera selección. En este caso los mejores resultados se obtienen con XBoost, con una accuracy de 0.71. Al analizar las características se puede observar cómo concentra la importancia en un número limitado de variables.

Estos resultados son coherentes con los obtenidos en el apartado anterior. Además, si se analiza el segundo mejor modelo en términos de accuracy, Adaboost, se puede observar este mismo comportamiento. La temperatura máxima de la artesa, la velocidad mínima y el argón en la línea de flotación son las únicas variables que emplea para la clasificación.

En cambio, el resto de los modelos, que ofrecen un peor desempeño, reparten la importancia entre un gran número de características. Esto parece apuntar, como en el caso anterior, a que el resto de modelos no están consiguiendo generalizar por estar capturando ruido presente en los datos.

Del mismo modo que en el caso abordado en el apartado 5.2.1, se ha ido reduciendo de forma iterativa el número de variables, empleando *SequentialFeatureSelector*. Tras esta reducción todos los modelos parecen apuntar a que las variables que tienen un efecto más relevante son *Temperatura Artesa MÁXIMA (°C)* y *Velocidad MÍNIMA (m/min)*. Otras variables que tienen un efecto en ciertos modelos son *Ar line flotación (l/min)*, *Temperatura SHC (°C)*, *ppm O2 EHC* o *Temperatura Artesa MEDIA (°C)*. Sin embargo, estas últimas fluctúan entre modelos.

Si se eliminan estas variables fluctuantes, es decir, se modeliza empleando únicamente la temperatura máxima de la artesa y la velocidad mínima, se obtienen resultados mejores, o como mínimo iguales, que empleando un número mayor de variables.

Los mejores resultados se obtienen con XGBoost, el modelo no se ve afectado negativamente en términos de accuracy, manteniendo un 0.71, esto es más del doble de lo que se esperaría por azar. Otros modelos que proporcionan un desempeño similar, pero inferior son un árbol de clasificación, Gradient Boosting o una red neuronal. El modelo basado en un árbol de decisión únicamente emplea la variable *Temperatura Artesa MÁXIMA (°C)* para realizar la clasificación. En cambio, Gradient Boosting y la red neuronal usan *Temperatura Artesa MÁXIMA (°C)* y *Velocidad MÍNIMA (m/min)*. En los tres casos se alcanza una accuracy de 0.69.

Los resultados obtenidos con los diferentes modelos se resumen en la siguiente tabla.

Modelo	Accuracy
Regresión Logística	0.49
KNN	0.59
Árbol de Clasificación	0.69
Random Forest	0.65
Gradient Boosting	0.69
XGboost	0.71
AdaBoost	0.65
SVM	0.61
Naive Bayes	0.65
Red neuronal	0.69

Cuadro 2: Accuracy para diferentes modelos de clasificación.

A continuación, se presentan la matriz de confusión para el modelo Xboost.

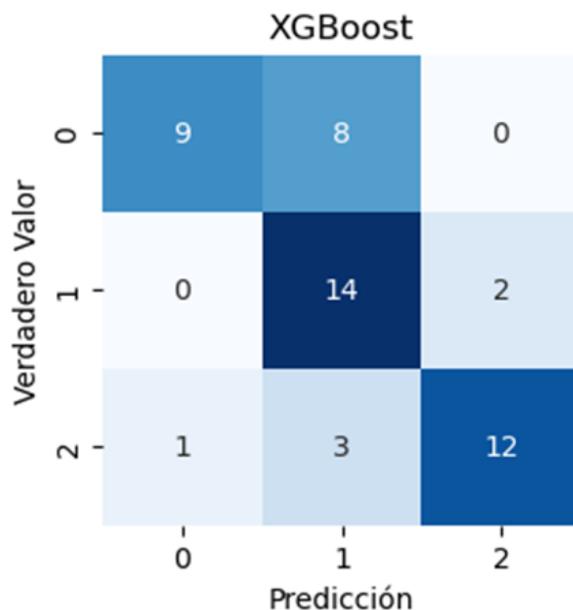


Figura 7: Matriz de confusión para clasificación en tres clases

En esta matriz de confusión se puede apreciar como el modelo está diferenciando entre la clase 0 y 2. Los errores que comete están asociados a clasificar incorrectamente en la clase 1 instancias de las otras clases.

Las características de interés son aquellas que generan un tamaño de inclusión que se sitúe en los valores extremos de la distribución, por tanto, los errores en la clasificación de las instancias que se sitúan en el segundo tercil es menos relevante.

Es interesante notar que la matriz de confusión del modelo de Gradient Boosting y la correspondiente al árbol de decisión coinciden, a pesar de que el árbol solo basa la clasificación en la variable 'Temperatura Máxima'. Esto puede deberse a que las variables 'Velocidad mínima' y 'Temperatura Máxima' estén altamente correlacionadas. La correlación que presentan es de un 0.58, esta correlación puede considerarse moderadamente alta y podría implicar que la información aportada por la variable Velocidad Mínima sea redundante.

Sin embargo, si se elimina la variable 'Velocidad Mínima' el modelo Gradient Boosting ve afectado negativamente su desempeño puede que este modelo esté empleando relaciones complejas entre las dos variables para corregir las predicciones.

A continuación, se presentan las PDP para el modelo XGBoost.

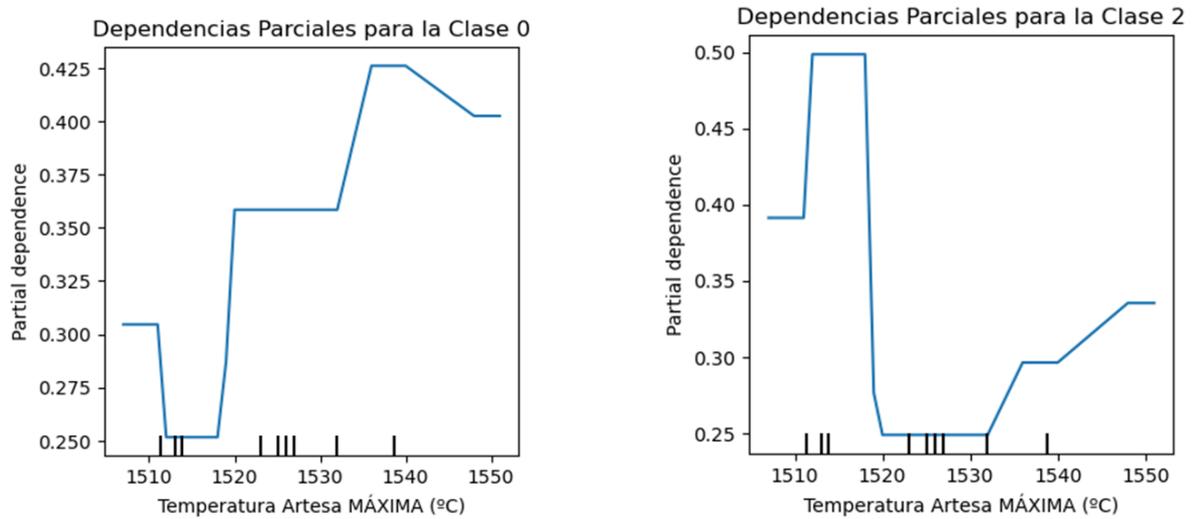


Figura 8: PDP para Temperatura Artesa MÁXIMA (°C)

En estas se puede apreciar como la probabilidad de que una instancia se clasifique en la clase 0 (menor tamaño de inclusión), crece a medida que aumenta la temperatura máxima de la artesa, tomando su valor máximo cuando esta se sitúa entre 1536 y 1540 °C. Para la clasificación en la clase 2 (inclusiones de mayor tamaño), las instancias se clasificaran en la en esta clase con una mayor probabilidad cuando la temperatura no supere los 1519 °C.

Es interesante notar que al inicio de ambas PDP existe una cresta inicial, esto parece indicar que para temperaturas más bajas que 1512°C el modelo puede estar cometiendo errores, a pesar de ello el hecho de que la probabilidad de pertenencia a la clase 2 sea mayor, para estos valores de temperatura, es consistente con los resultados obtenidos en otros modelos.

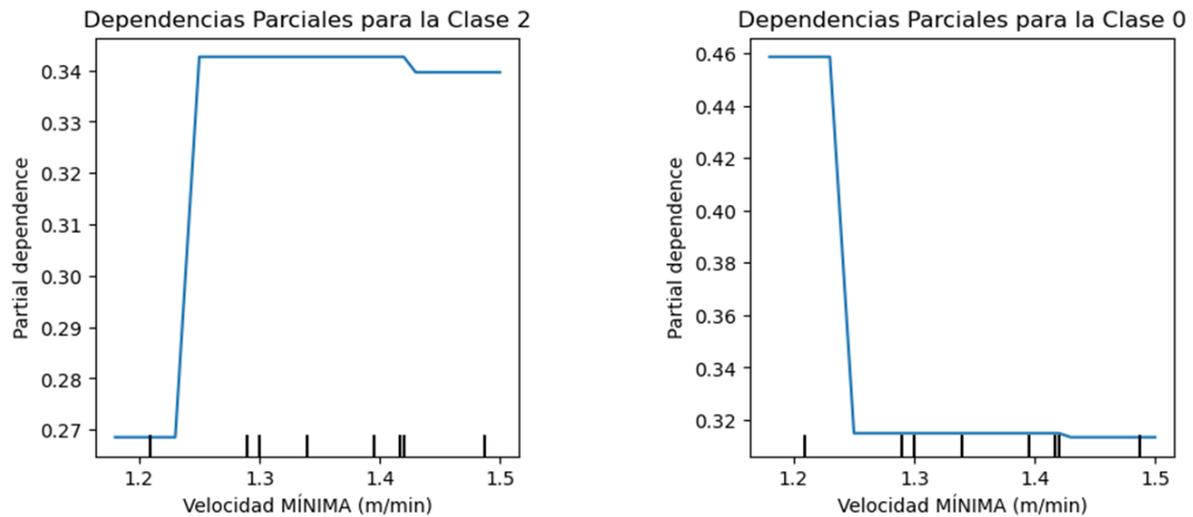


Figura 9: PDP para Velocidad MÍNIMA (m/min)

En el caso de la velocidad mínima, la probabilidad de que la instancia sea clasificada en la clase 0 disminuye abruptamente a partir de una velocidad de 1.25 m/min.

Respecto a la clasificación en la clase 2, el punto de inflexión se produce en 1.25 m/min, velocidad a partir de la cual la probabilidad de ser clasificado en la clase 2 disminuye.

Por tanto, al igual que en el apartado anterior se llega a la conclusión de que a mayor temperatura máxima en la artesa menor será el tamaño de inclusión, por otro lado, cuanto mayor sea la velocidad mínima, mayor será el tamaño.

Para analizar en mayor profundidad los resultados obtenidos en las PDP del modelo XGBoost, se ha representado también las PDP para el modelo Gradient Boosting.

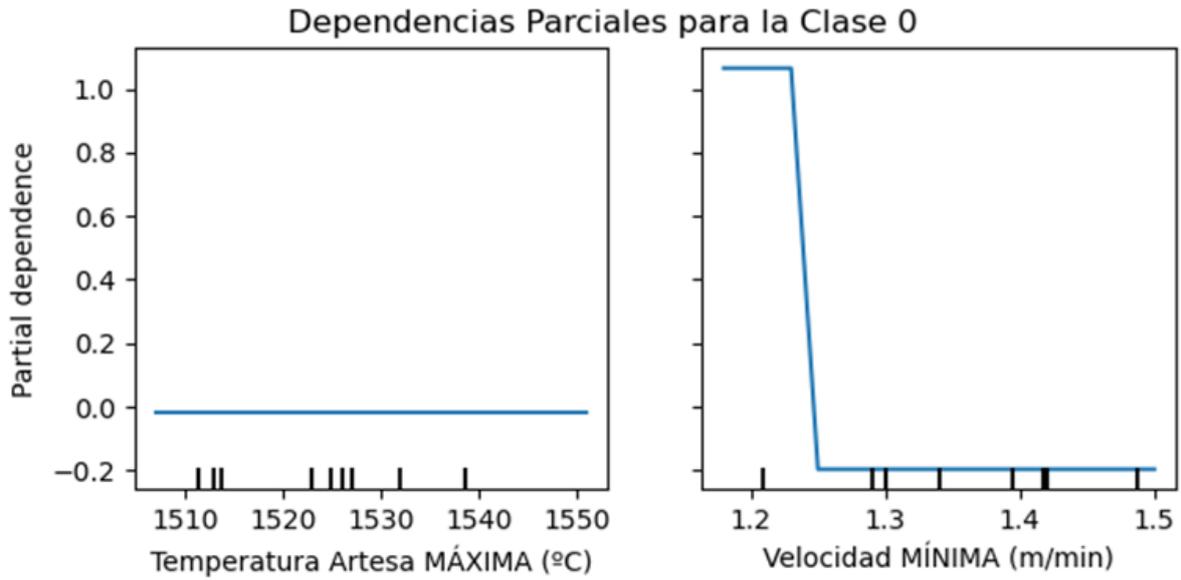


Figura 10: PDP para la clasificación en la clase 0

Este modelo parece estar empleando la velocidad mínima únicamente para la clasificación en la clase 0, la variable *Temperatura Artesa MÁXIMA (°C)*, en cambio, permanece constante. Parece que esta última variable no juega un papel relevante en la clasificación de la clase 0. Sin embargo, la probabilidad de ser clasificado en la clase 0 disminuye a medida que la velocidad aumenta. Concretamente parece el punto de inflexión parece estar en la 1.23 m/min, a partir de este valor el modelo deja de clasificar en la clase 0.

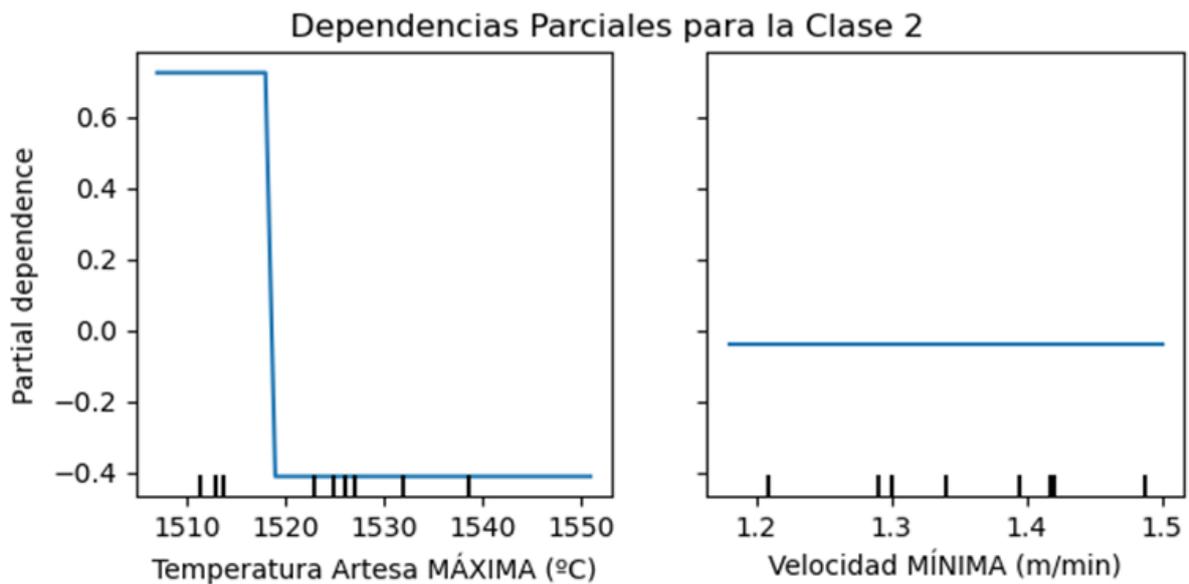


Figura 11: PDP para la clasificación en la clase 2

El efecto contrario se observa en las PDP de la predicción de la clase 2 (mayor tamaño de inclusión). Mientras que la PDP de la velocidad permanece constante, la asociada a la temperatura fluctúa reduciéndose la probabilidad de que la observación sea clasificada en esta clase a medida que aumenta la temperatura. Esta fluctuación se produce para una temperatura de 1518°C, es decir, a partir de esta temperatura el modelo ya no clasifica la observación en la clase 2.

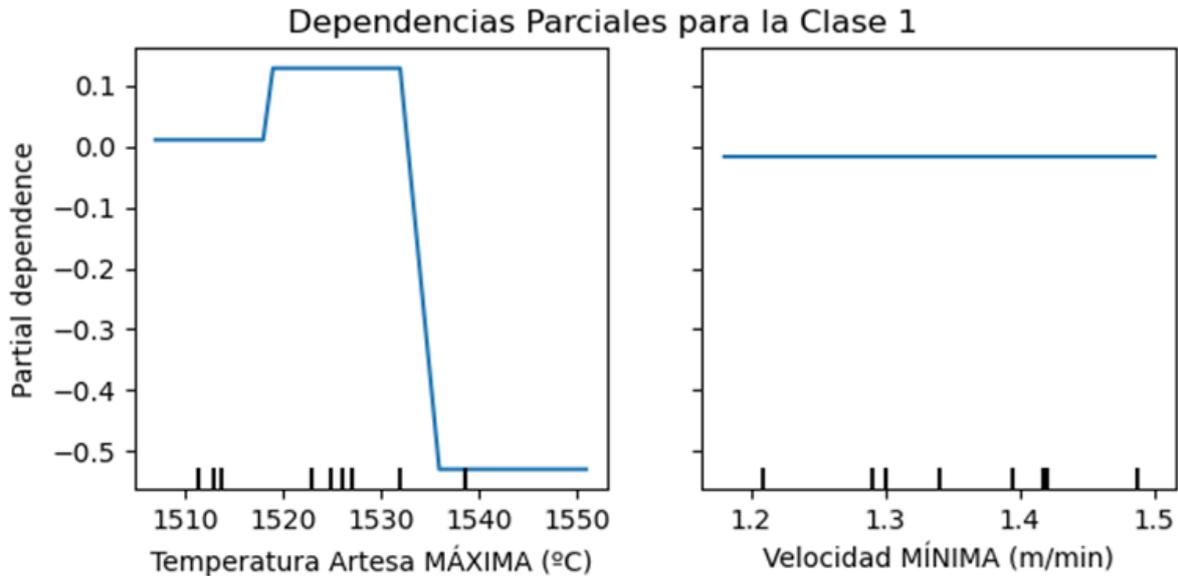


Figura 12: PDP para la clasificación en la clase 1

En cuanto a las PDP asociadas a la clase 1 (tamaño de inclusión intermedio), únicamente fluctúan con la temperatura, sin embargo, el comportamiento es más complejo que en el caso anterior. En primer lugar, el gráfico es estable hasta alcanzar una temperatura de 1518°C, aquí experimenta un incremento hasta 1532°C, temperatura a partir de la cual la dependencia parcial cae abruptamente, por tanto, el modelo deja de clasificar las instancias en la clase 1.

Estos resultados son coherentes con los obtenidos en la clasificación para las clases 0 y 2, ya que cuando la temperatura se sitúa por debajo de 1518°C la instancia se clasifica en la clase 2, con lo que la probabilidad de ser clasificada en la clase 1 disminuye. Lo mismo ocurre con la clasificación en la clase 0.

Por tanto, se reafirman las conclusiones obtenidas en el apartado anterior. En ambos casos, los valores de temperatura y velocidad que producen la fluctuación en la clasifi-

cación son coincidentes. Las PDP de segundo apartado, aportan información sobre que variable interviene en mayor medida en la clasificación de cada clase.

7 Conclusiones.

Resumen del Alcance y Resultados del Proyecto

El objetivo principal de este trabajo radicaba en estudiar los factores responsables de la aparición de inclusiones no metálicas durante el proceso de producción del acero. Esto se ha hecho mediante el empleo de técnicas de minería de datos y aprendizaje automático. Las clasificaciones aplicadas fueron de dos tipos: una clasificación binaria donde la variable objetivo era 0 o 1 según si el tamaño de inclusión promedio estaba por debajo o por encima del valor mediano, y una clasificación ternaria determinada por los terciles del tamaño de inclusión.

Los resultados más exitosos se lograron con los modelos XGBoost y Gradient Boosting. El modelo Gradient Boosting obtuvo una *accuracy* de 0,81 en clasificación binaria, y XGBoost logró una *accuracy* de 0,71 en clasificación ternaria. Se encontró que las variables críticas para la clasificación eran la temperatura máxima de la artesa junto con la velocidad mínima del acero durante la colada.

Discusión

A la vista de los resultados obtenidos en el conjunto de modelos, parece quedar patente que las variables que desempeñan un papel más relevante son la temperatura máxima de la artesa y la velocidad mínima. Los algoritmos empleados coinciden en dotar a estas variables de una importancia clave para la clasificación, tanto binaria como ternaria.

- **Temperatura Máxima de la Artesa:** Según los resultados obtenidos existe una relación inversa entre la temperatura máxima de la artesa y el tamaño promedio de las inclusiones. A mayor temperatura máxima en la artesa menor tamaño de inclusión. El mecanismo a través del cual se produce este efecto no es claro, sin

embargo, estudios como el de Yuji Miki y B.G. Thomas (1999) abordan la influencia que tiene la temperatura de la artesa en la eliminación de las inclusiones a través de la boyancia térmica. Este proceso consiste en la variación en el flujo de acero dentro de la artesa a causa de la disminución en la densidad del acero. Según estos autores, esto puede favorecer que las inclusiones floten con mayor facilidad a la escoria y que en consecuencia se favorezca su eliminación.

Otro mecanismo a través del cual la temperatura máxima de la artesa podría estar jugando un papel relevante en la reducción del tamaño de inclusión, es a través del aumento de la solubilidad del acero[17]. Este aumento en la solubilidad podría generar que los elementos causantes de las inclusiones se dispersen en el acero líquido[18].

Sin embargo, existen mecanismos adicionales que pueden generar los efectos contrarios, como por ejemplo el mayor desgaste del refractario de la artesa[19].

- **Velocidad Mínima:** Se observó una relación positiva entre la velocidad mínima y el tamaño de inclusión, esto es, a mayor velocidad, mayor tamaño de inclusión. Los procesos que pueden estar dando origen a este resultado son complejos y plantean posibles efectos ambivalentes. Posibles justificaciones para este patrón incluyen un mayor desgaste del refractario debido a la mayor velocidad del acero, la posibilidad de que un menor tiempo de enfriado impida que las inclusiones migren hacia la superficie del metal, o el aumento en la probabilidad de colisión entre inclusiones a causa de la mayor velocidad de flujo [20]. Sin embargo, la evaluación de las causas de la influencia de esta variable pueden ser complejas y requerirían de un análisis exhaustivo.

Es destacable notar que, en este proceso, las variables velocidad mínima y temperatura máxima están intrínsecamente relacionadas, dado que la temperatura del acero afectará directamente a su fluidez y en consecuencia a la velocidad a la que se desplazará el acero líquido. Esto podría explicar por qué algunos modelos no empeoran tras la eliminación de la variable "Velocidad mínima". Estos modelos podrían estar capturando la suficiente información de la variable temperatura máxima y no requerir de la información adicional que aporta la variable velocidad.

Limitaciones y Futuras Investigaciones

Aunque los resultados son prometedores y coherentes entre los modelos, pueden estar sujetos a múltiples limitaciones debido al reducido tamaño del dataset. Es crucial ampliar el número de coladas analizadas para poder validar estos resultados. La realización de un mayor número de ensayos ayudaría a reducir el ruido y a mejorar la precisión de los modelos predictivos. Además, un número mayor de muestras permitiría revelar la existencia de variables adicionales que puedan estar desempeñando un papel relevante pero cuyo efecto no se puede cuantificar debido a la pérdida en la precisión que sufren los modelos al aumentar el número de variables.

Para posteriores análisis, sería interesante incorporar información relativa a la composición de las inclusiones detectadas. La sinergia entre esta información y las variables de producción podría ayudar a identificar las etapas del proceso productivo en las que se forma cada tipo de inclusión. Esto permitiría expandir la comprensión de las interacciones entre los materiales y los procedimientos empleados en la producción de acero, facilitando de este modo la determinación de las causas subyacentes a la formación de las inclusiones.

Referencias

- [1] World Steel Association. World steel in figures 2024, 2024. Accessed: 2024-06-06.
- [2] Y. Murakami, S. Kodama, and S. Konuma. Quantitative evaluation of effects of non-metallic inclusions on fatigue strength of high strength steels. i: Basic fatigue mechanism and evaluation of correlation between the fatigue fracture stress and the size and location of non-metallic inclusions. *International Journal of Fatigue*, 11(5):291–298, 1989.
- [3] H. Tervo, A. Kaijalainen, T. Pikkarainen, S. Mehtonen, and D. Porter. Effect of impurity level and inclusions on the ductility and toughness of an ultra-high-strength steel. *Materials Science and Engineering: A*, 697:184–193, 2017.
- [4] T. V. Shibaeva, V. K. Laurinavichyute, G. A. Tsirlina, A. M. Arsenkin, and K. V. Grigorovich. The effect of microstructure and non-metallic inclusions on corrosion behavior of low carbon steel in chloride containing solutions. *Corrosion Science*, 80:299–308, 2014. Influence of inclusions on initiation of pitting corrosion and stress corrosion cracking of X70 steel in near-neutral pH environment.
- [5] D. Krewerth, T. Lippmann, A. Weidner, and H. Biermann. Influence of non-metallic inclusions on fatigue life in the very high cycle fatigue regime. *International Journal of Fatigue*, 84:40–52, 2016.
- [6] A. B. Arnaldo, C. B. Juan, and H. Alberto. Acondicionamiento térmico de la fundición en antecrisol con inyección de oxígeno. *Fundidores*, 2006.
- [7] K Walker. The basics of iron and steelmaking. *Steel Times International*, 36(3):38, 2012.
- [8] C S Russell and W J Vaughn. *Steel production: processes, products, and residuals*. RFF Press, 2013.
- [9] P Kaushik, J Lehmann, and M Nadif. State of the art in control of inclusions, their characterization, and future requirements. *Metallurgical and Materials Transactions B*, 43(4):710–725, 2012.
- [10] H V Atkinson and G Shi. Characterization of inclusions in clean steels: a review in-

- cluding the statistics of extremes methods. *Progress in Materials Science*, 48(5):457–520, 2003.
- [11] L J House and R Rungta. A review of advances in noninvasive methods for detection and characterization of inclusions. In *Proceedings of a symposium held in conjunction with the 1988 World Materials Congress on "Inclusions and their Influence on Material Behavior"*, pages 93–108, Illinois, September 1988.
- [12] Estela Ruiz Martínez. *Optimización del proceso de fabricación de acero en horno de arco eléctrico mediante algoritmos machine learning y procedimientos experimentales específicos para la detección de macroinclusiones no metálicas*. PhD thesis, Universidad de Cantabria, 2021.
- [13] Jerome H Friedman. Greedy function approximation: A gradient boosting machine. *Annals of statistics*, pages 1189–1232, 2001.
- [14] Scott M Lundberg and Su-In Lee. A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, 2017.
- [15] Christoph Molnar. *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Leanpub, 2020.
- [16] Yuji Miki and B.G. Thomas. Modeling of inclusion removal in a tundish. *Metall Mater Trans B*, 30:639–654, 1999.
- [17] M. C. Flemings. (b) formation of oxide inclusions during solidification. *International Metals Reviews*, 22(1):201–207, 1977.
- [18] Non-metallic inclusions in steel. https://www.machinemfg.com/non-metallic-inclusions-in-steel/?utm_content=cmp-true. Accessed: 11 June 2024.
- [19] J L Enríquez Berciano, E Tremps Guerra, S D Elío de Bengy, and D Fernández Segovia. *Colada de acero*. 2009.
- [20] L. Zhang, S. Taniguchi, and K. Cai. Fluid flow and inclusion removal in continuous casting tundish. *Metallurgical and Materials Transactions B*, 31:253–266, 2000.