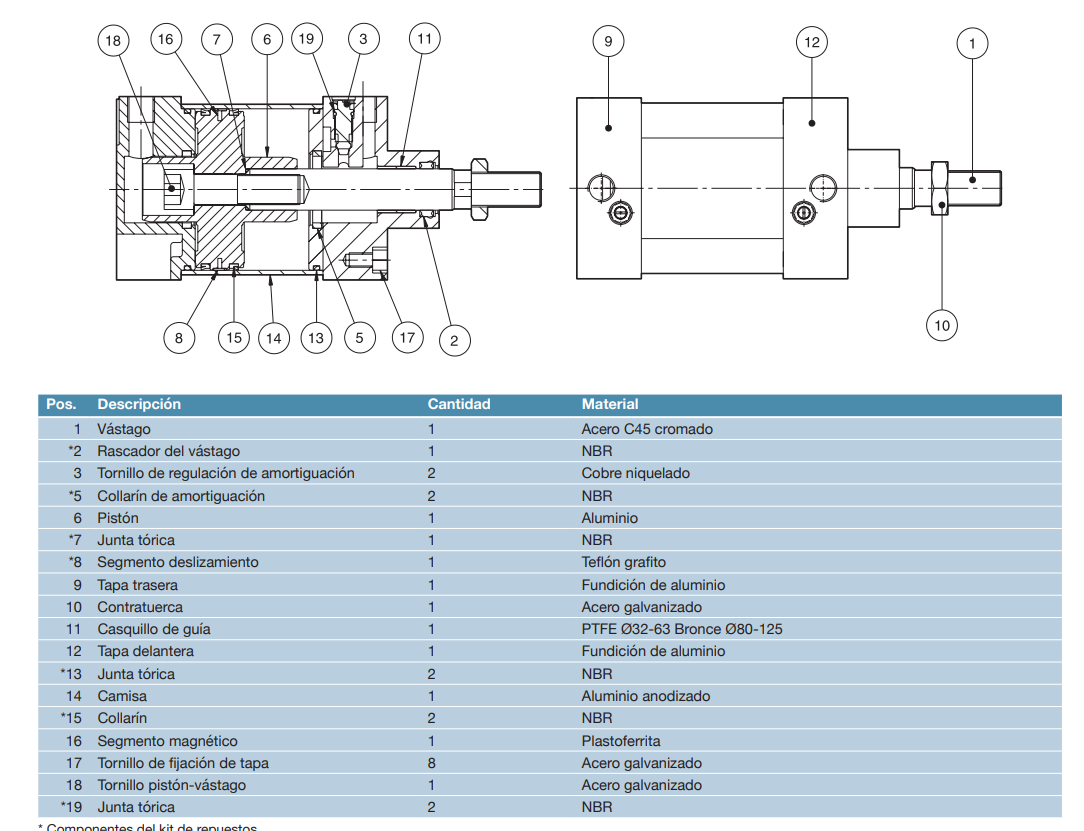
**CALCULO PREDICTIVO DE CILINDROS ESPECIALES**

MEMORIA TRABAJO DE FIN DE MASTER

1. INTRODUCCION

En este proyecto vamos a crear un modelo predictivo para el cálculo de presupuestos para cilindros neumáticos especiales. Para que sea mas entendible, vamos a empezar explicando que es un cilindro y con que partes está compuesto.

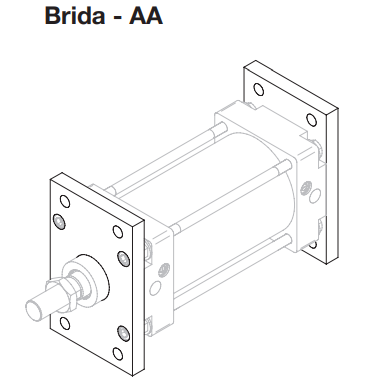
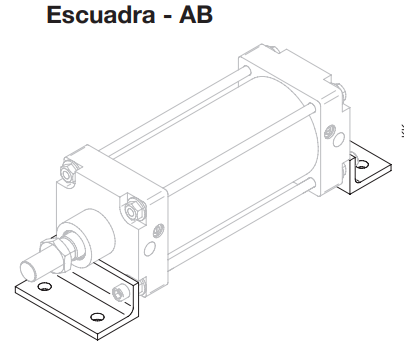
Un cilindro neumático o actuador neumático utiliza el aire presurizado como fluido para generar un movimiento linear. En la siguiente imagen se pueden apreciar las partes de un actuador estándar.

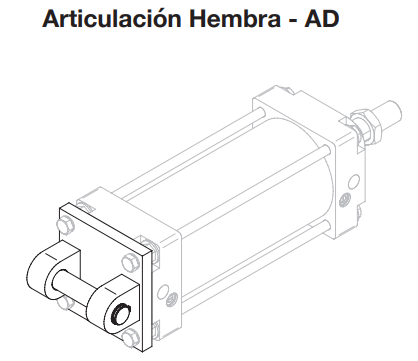
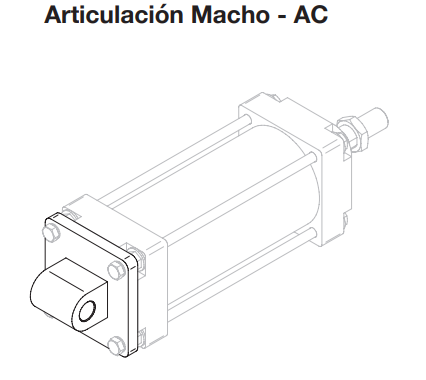


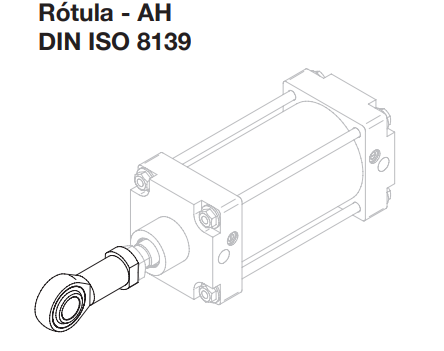
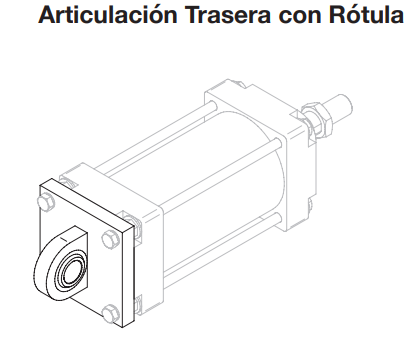
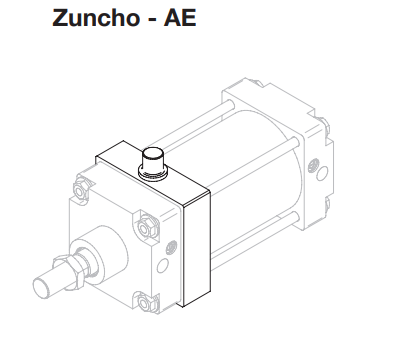
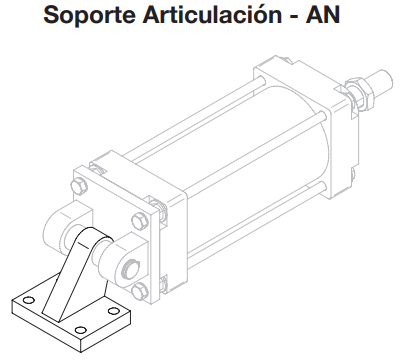
El aire comprimido se introduce desde las tapas delantera y trasera (9 y 12). Con este aire se genera el movimiento lineal del pistón (6) y del vástago (1) que va amarrado al pistón. Este movimiento puede mover distintas cantidades de peso según el tamaño del pistón y de la presión del aire introducido.

Los cilindros especiales se diferencian de lo anterior en que las medidas y los materiales pueden ser personalizados a las peticiones del cliente. La razón de esto es porque a veces los clientes necesitan cilindros mas compactos por falta de espacio, mas resistentes para aplicaciones muy pesadas o mas duraderos para ambientes corrosivos entre otros.

A estos cilindros base, habitualmente se le suele instalar distintos amarres para facilitar la adaptación al sistema del cliente, estos son algunos de ellos:







1. MOTIVACIÓN

La motivación de este proyecto ha por la necesidad de ahorra tiempo en la preparación de presupuestos. Se ha detectado que se pierde mucho tiempo recalculando cilindros muy parecidos pero que se diferencian en algunos detalles y también en cilindros que suelen salir de lo estándar ya que se suelen recalcular y revisar todas las partes del cilindro y su fabricación.

El objetivo es crear un programa para utilizar en los momentos de ajetreo para calcular el presupuesto del cilindro en un instante.

1. OBTENCION DE DATOS

Los datos se han conseguido de la empresa donde actualmente trabajo, se han descargado varias tablas que contienen los datos que nos interesan a archivos .csv para poder tratarlos:

* ‘ARTICULOS.csv’: es la tabla donde recoge las referencias de los artículos de la empresa junto con las descripciones, entre otros datos.
* ‘CONJUNTO FAV.csv’: contiene los conjuntos con su respectiva sequencia, los conjuntos son aquellos artículos que se fabrican usando otros artículos. Un articulo esta generado por varios artículos, por ejemplo un cilindro.
* ‘DET\_CONJFAV.csv’: aquí se definen los conjuntos anteriormente mencionados, está compuesto por las secuencias de los conjuntos de la tabla ‘CONJUNTO FAV.csv’ y las secuencias de los artículos de la tabla ‘ARTICULOS.csv’.
* ‘MOVIMIENTO ALMACEN EXCEL.xslx’: los artículos tienen movimientos de salida y de entrada en el inventario, esta tabla registra los tipos de salida que ha tenido junto con la fecha exacta. Los tipos de movimiento pueden ser: compra, venta, fabricación, regularización de existencias…
* ‘HISTORICO ALVARANES COMPRA.csv’: aquí se registra la información de las compras a proveedores, junto con las referencias de los artículos que se han comprado.

1. TRATAMIENTOS DE DATOS

El primer problema a tratar ha sido que los datos que nos interesaban estaban distribuidos en distintas tablas y además utilizando secuencias para relacionar estas.

El segundo y gran problema que nos hemos encontrado que a lo largo de los años y distintas personas que han pasado por la empresa, la codificación de los artículos ha ido cambiando y no se ha respetado demasiado. Debido a esto, y después de observar que las descripciones si se han mantenido con el tiempo, decidí construir un dataframe con las siguientes variables sacadas de las descripciones.

Las variables los he elegido personalmente debido a mi conocimiento en la industria ya se cuales influyen en el cálculo.

MATERIALES DE ELEMENTOS PRINCIPALES

* MP CAMISA: materia prima de la camisa.
* MP VASTAGO: materia prima del vástago.
* MP VARILLAS: materia prima de las varillas.
* MP CABEZA DE PISTON: materia prima del pistón.
* MP TAPA GUIA: materia prima de la tapa guía.
* MP TAPA TRASERA: materia prima de la tapa trasera.

DIMENSIONALES

* DIAMETRO: es el diámetro de la camisa, esta medida es la medida principal en dos dimensiones, todo el cilindro se dimensiona respecto a esta medida. Ya que en cuanto mayor es el diámetro mayor fuerza debe soportar el cilindro.
* CARRERA: la carrera es el recorrido que hace el cilindro, cuanto mas larga es la carrera mas largo es el cilindro, por lo tanto conseguimos la dimensionalidad del cilindro a lo largo, que es el que el que nos faltaba.

AMARRES Y EXTRAS

* AMARRE 0,1,2,3…: estos son los amarres que he explicado antes, como un cilindro puede tener mas de uno la cantidad de columnas lo define el que mas tiene.
* EXTRA 0,1,2,3…: como hay cilindros muy distintos, algunos con doble vástago, otros son rotativos que llevan piñones y cremalleras, algunos llevan una electroválvula montada… se añaden los extras mas comunes en las variables.

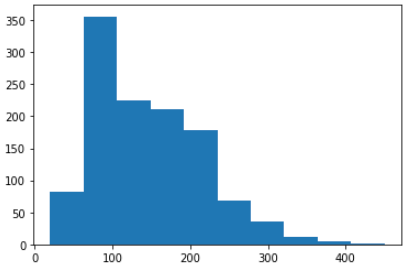
OTROS

* KIT: el kit esta compuesto por las tapas, pistón y juntas y es lo que se utiliza en los cilindros estándares. Algunos cilindros especiales utilizan los kits ya que lo que hace el cilindro especial puede ser otro elemento. Como estos kit son tan baratos ya que se fabrican en series, influye mucho en el precio final, es por eso por lo que se ha elegido como una variable.
* AMORTIGUACION: la amortiguación en los cilindros se utiliza para que el pistón no golpee fuertemente las tapas al llegar al final de su recorrido. Así se evitan vibraciones que puedan estropear el sistema al paso del tiempo. En algunas aplicaciones no suele ser necesario y en otras si, incorporar este sistema influye al cilindro añadir mas piezas por lo tanto significa mayor coste.
* ALTA TEMPERATURA: los cilindros que deben sufrir altas temperaturas ambientes, los que van instalados en hornos, por ejemplo, necesitan unas juntas especiales que lo resistan. Hay una diferencia importante en el coste de estas juntas respecto a las estándares.
* UNIDADES: como en la fabricación influye mucho que se haga una unidad o varios se incluye la variable de cuantos cilindros se han fabricado a la vez.

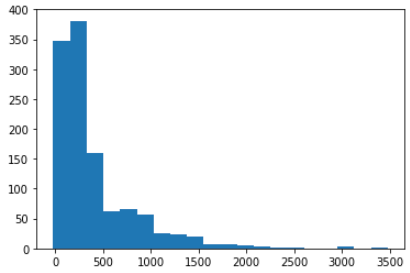
EXPLORACION DE DATOS

Una vez que tenemos el dataframes con los datos necesarios vamos a explorar lo que hemos conseguido para ver si tiene sentido:

Lo primero vamos a analizar la distribución de los datos respecto al diámetro:

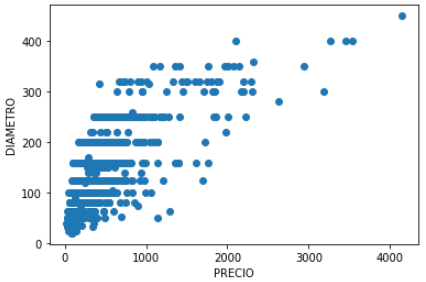


También respecto a la carrera:



Se ve que la mayor concentración de cilindros que tenemos son entre los diámetros 80-160 mm y carreras 0-500 mm, estos son los cilindros mas habituales, es de esperar que nuestro modelo prediga mejor cilindros de estas medidas.

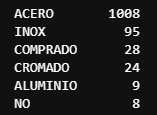
Por otra parte vamos a ver la influencia de estas variables en los precios:



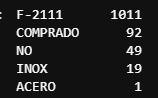
Se puede apreciar que hay una relación clara, cuanto mayor es el diámetro mayor es el precio, por lo tanto vemos que los datos conseguidos tienen sentido.

Por otra parte hemos analizado las materias primas para saber cuales son las mas comunes.

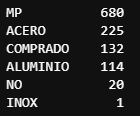
Vástago



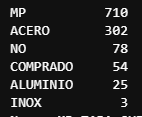
Varillas



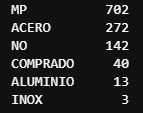
Cabeza de Pistón



Tapa guía



Tapa trasera



El primer problema que se detecta es que las muestras son muy pocas, en los casos de los materiales pocos comunes puede haber bastante desviación en la predicción por los pocos datos de entrenamiento. Por ejemplo, se ve claro en el caso de la cabeza de pistón inoxidable, solo hay 1 muestra con lo que no podemos esperar una buena predicción en este caso.

También se ve que hay algunos “NO” los cuales son los NaN convertidos en variable categórica. En algunos casos los cilindros pueden no llevar algún elemento y tener duplicado otro tal y como se ha explicado anteriormente. Es por esto por lo que como norma general tomo los NaN como si no llevase esta pieza.

LIMPIEZA DE DATOS

PREPARACION DEL MODELO

Se han seleccionado los algoritmos predictivos mas populares de aprendizaje supervisado:

* Decision Tree Regressor
* KNeighbors Regressor
* Linear Regression
* Random Forest Regressor

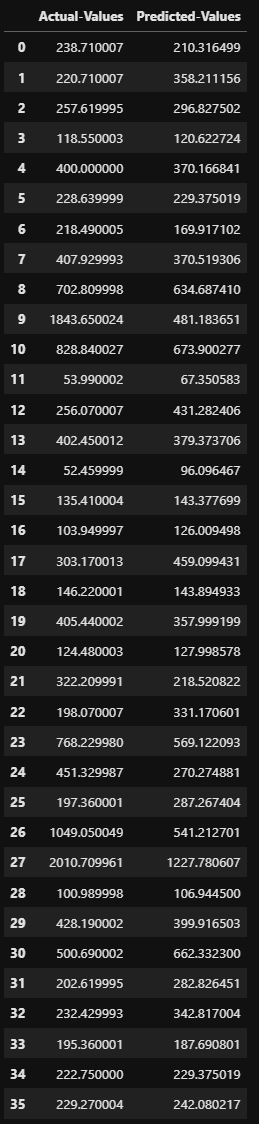
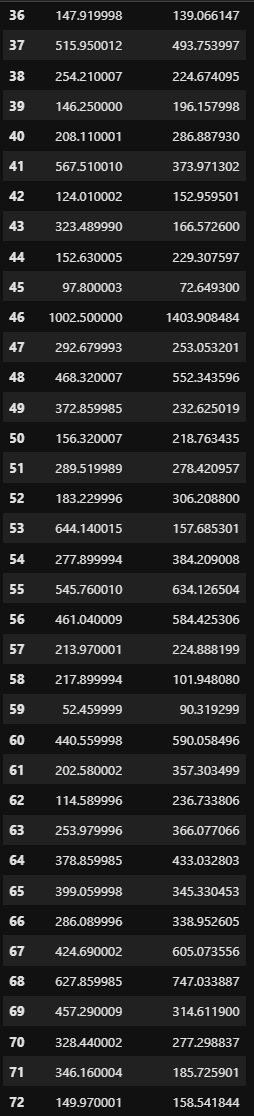
Para la evaluación del modelo se ha definido una función que nos devuelve las siguientes métricas:

* Mean Squared Error
* Mean Absolute Error
* Mean Absolute Percentage Error
* Explained variation

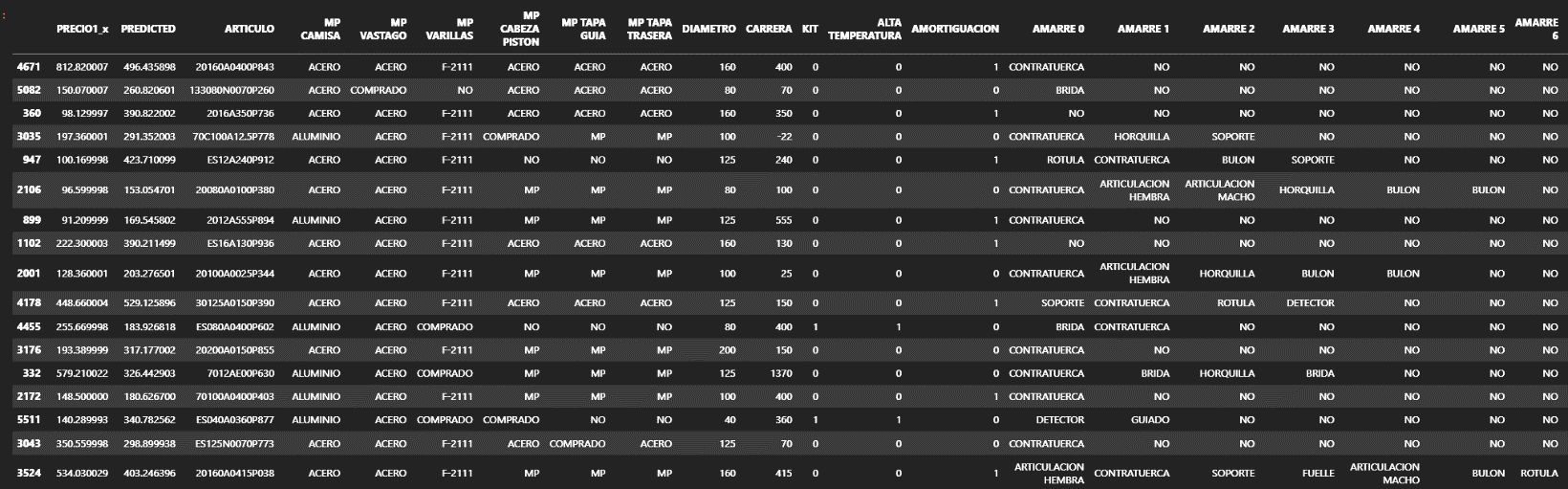
Creamos distintos pipelines con los modelos seleccionados y distintos parámetros para probar cual es el que mejor predice de todos y el que mejores métricas nos dan con el “Random Forest Regressor”.

Esto se debe a que tenemos muchas variables categóricas y pocas numéricas, por lo tanto los modelos que se basan en cálculos funcionan peor. Sin embargo, como el Random Forest se basa en prueba y error al final recoge mayor cantidad de posibilidades.

En la siguiente tabla se pueden apreciar algunos de los resultados obtenidos comparados con los valores reales, se aprecia que en algunos casos el modelo funciona muy bien, sin embargo en otros hay una desviación muy grande.

Esto puede ser debido a lo que se ha comentado antes, hay poca diversificación en los materiales, se han analizado cuales son los que mas error tienen respecto al valor original.



Lo primero que llama la atención es que los cilindros que tienen muchos amarres sufren una gran desviación en la predicción.

Lo mismo pasa con los cilindros muy especiales que llevan 2 vástagos o llevan una electroválvula por ejemplo,

CONCLUSIONES

La conclusión principal después de haber analizado los resultados es que para mejorar el modelo predictivo habría que añadir mas datos, ya que se puede apreciar que en los casos menos comunes, los cuales he mencionado anteriormente, el modelo no funciona correctamente. Esto es seguramente porque habrá overfitting.