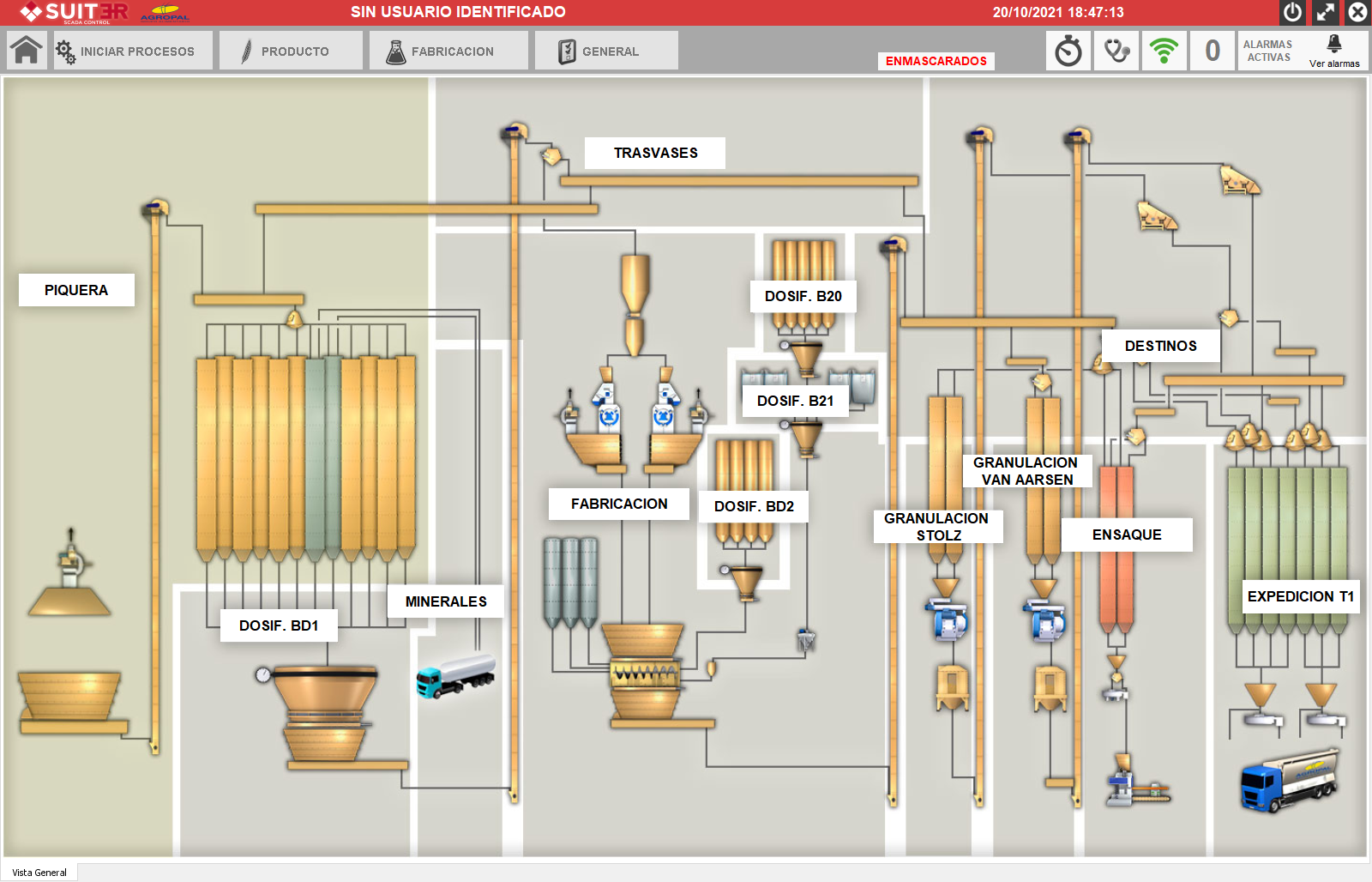
Trabajo Fin de Máster

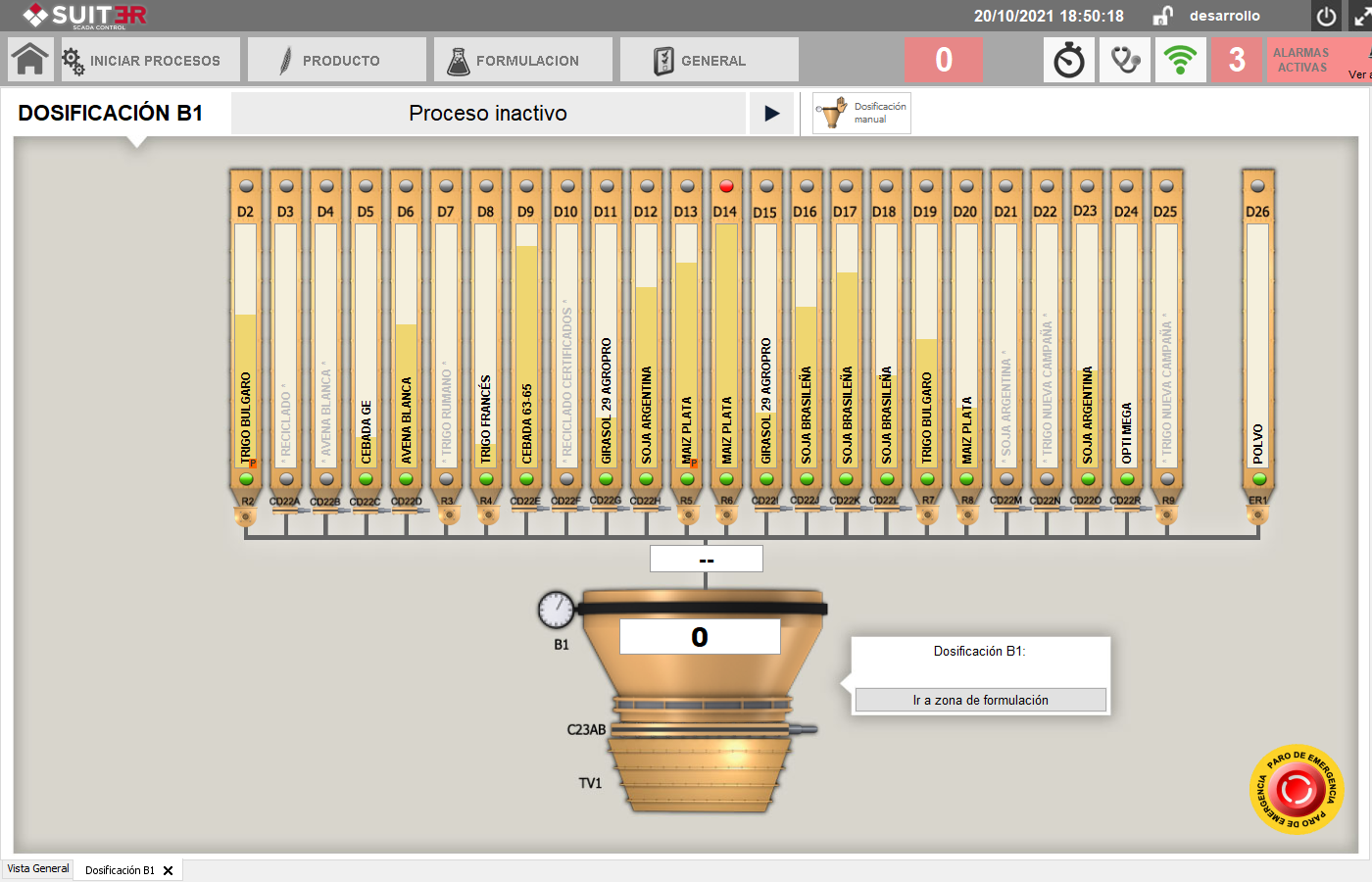
Máster en Ciencia de Datos e Ingeniería de Datos en la Nube

Predicción de duración de dosificación para corrección de colas en procesos automáticos de fabricación de pienso.

Auto: Jose Luis Casado Valero

Tutor: Luis de la Ossa Jiménez



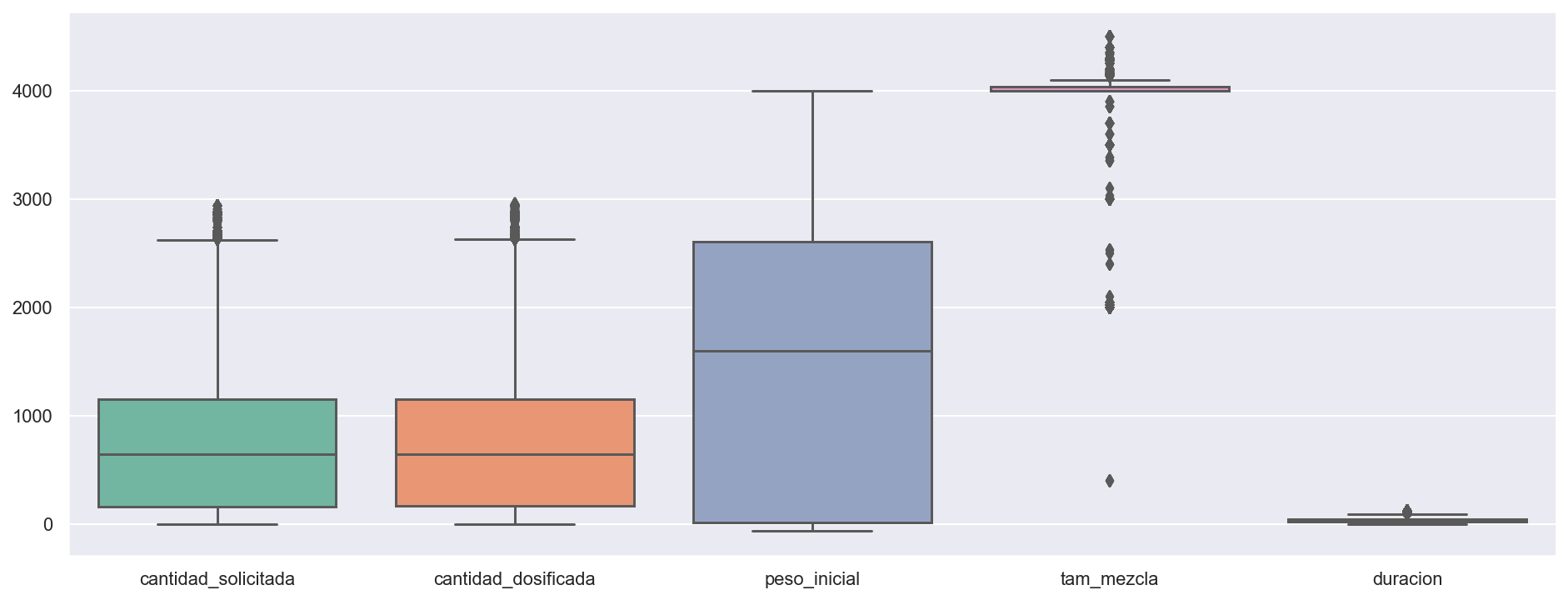


Meter descripción de los datos de los que disponemos

|  |  |  |
| --- | --- | --- |
| Nombre columna | Ejemplo | Descripción |
| Fecha\_inicio | 2013-08-14 16:51:48 | Fecha de inicio de la dosificación |
| Fecha\_fin | 2013-08-14 16:52:39 | Fecha de fin de la dosificación |
| Cantidad\_solicitada | 235.026 | Cantidad objetivo a dosificar |
| Cantidad\_dosificada | 236.5 | Cantidad real dosificada al finalizar |
| Mezcla | 1 | Nº de mezcla dentro de la fabricación |
| Pesada | 0 | Nº de pesada dentro de la mezcla |
| Peso\_inicial | 0.0 | Peso que marca la báscula antes de iniciar la dosificación |
| Manual | 0 | Booleano que indica si la dosificación ha sido manual o automática |
| Id\_lote\_destino | 232527 | Lote de fabricación. |
| Materia\_origen | 13 | Código de la materia origen. |
| Materia\_destino | 66 | Código de la materia destino |
| Id\_silo | 90 | Silo origen del que se dosifica |
| Tipo\_materia | Prima | Tipo de materia origen |
| Tipo\_destino | Premezcla | Tipo de materia destino |
| Densidad | 1.0 | Densidad de la materia origen |
| Tam\_mezcla | 500 | Suma teórica total de los ingredientes al finalizar todas las dosificaciones |
| Desviacion | 1.47 | Error real en la dosificación |

Enfoque del problema y

Creación de la columna duración que es la que vamos a predecir



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | C. Solicitada | C. Dosificada | P. Inicial | T. mezcla | Duración | Desviación | |
| Count | 395477 | 395477 | 395477 | 395477 | 395477 | | 395477 |
| Mean | 801.251 | 800.430 | 1342.961 | 4000.817 | 37.613 | | -0.8203 |
| Std | 744.934 | 745.126 | 1249.587 | 211.905 | 37.362 | | 30.4486 |
| Min | 0.000 | 0.000 | -64.000 | 400.000 | 0.000 | | -2623.09 |
| 25% | 167.984 | 168.000 | 11.000 | 4000.000 | 16.000 | | -3.000 |
| 50% | 668.040 | 664.000 | 1574.000 | 4000.000 | 25.000 | | 0.0040 |
| 75% | 1163.096 | 1168.000 | 2600.000 | 4040.000 | 49.000 | | 3.0198 |
| Max | 3121.984 | 3127.000 | 3995.000 | 4500.000 | 7786.000 | | 1179.891 |

Filtrado de datos:

* Cantidad solicitada mayor de 0.
* Cantidad solicitada menor de 1300.
* Cantidad dosificada mayor de 1.
* Cantidad dosificada menor de 1400.
* Desviación > -100.
* Desviación < 100.
* Peso inicial de la báscula menor de 2700.
* Duración menor de 120 segundos.

Como resultado nos queda una base de datos con 235.045 registros.

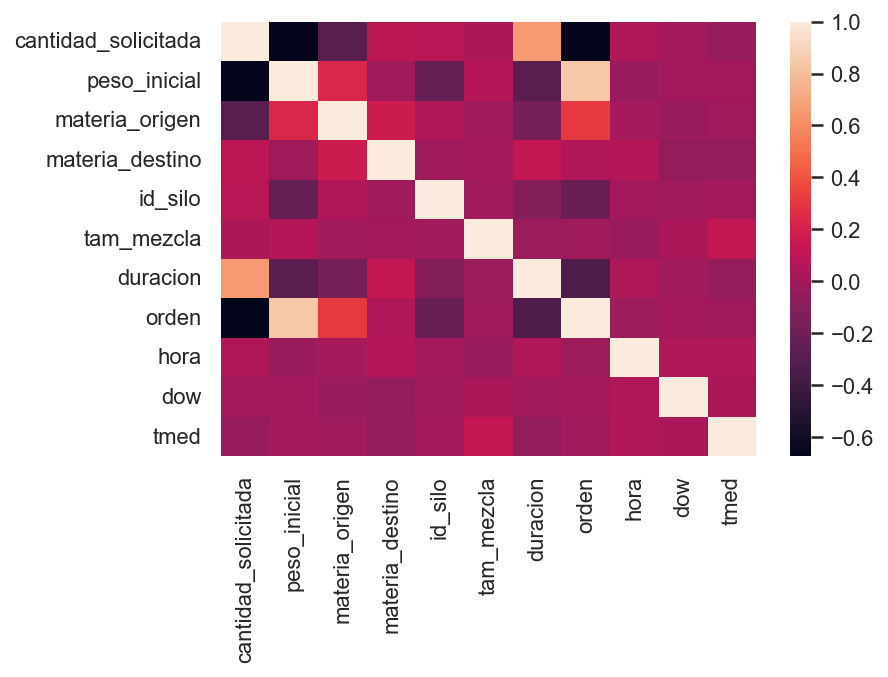
Columnas a eliminar:

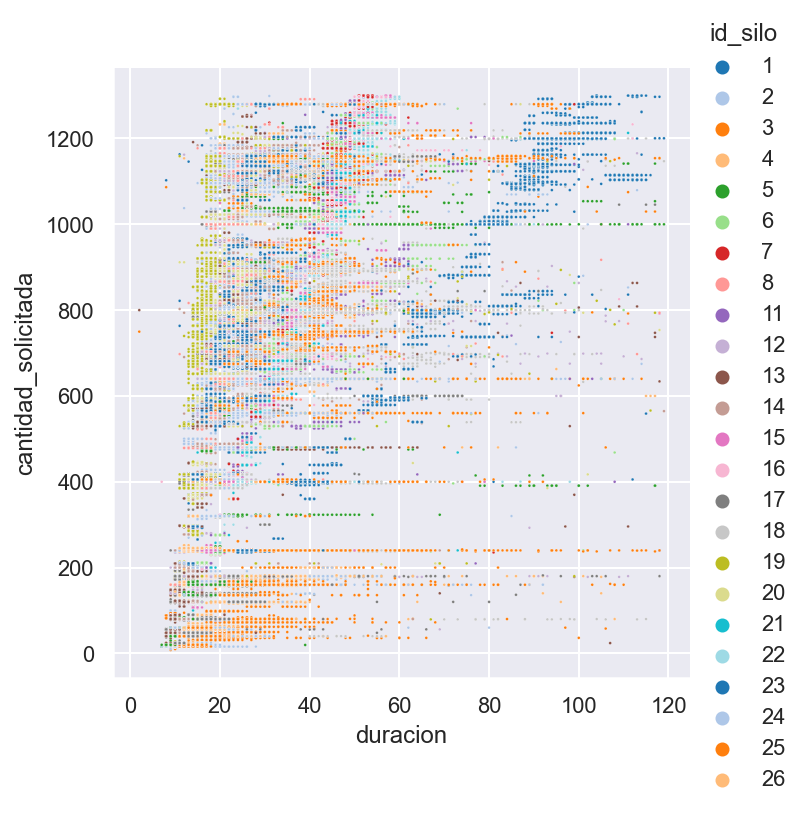
* Fecha\_fin.
* Cantidad\_dosificada.
* Mezcla.
* Pesada.
* Manual.
* Densidad.
* Tipo\_materia.
* Tipo\_destino.
* Desviación.

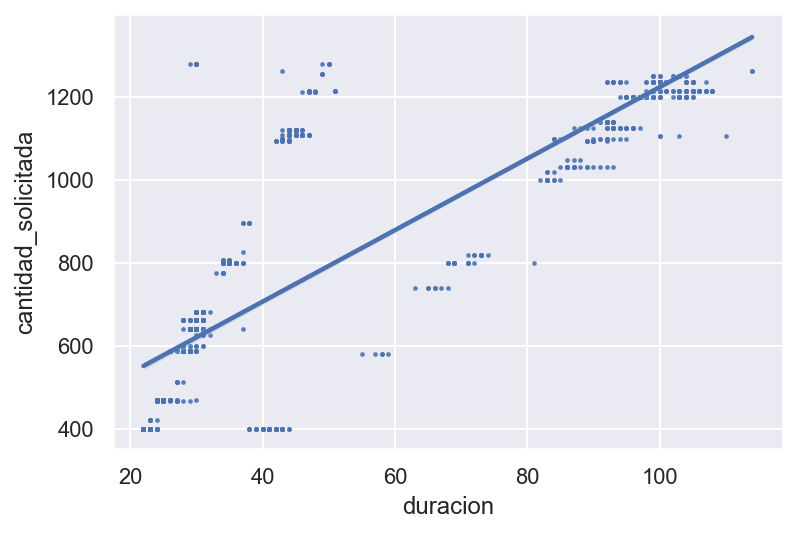
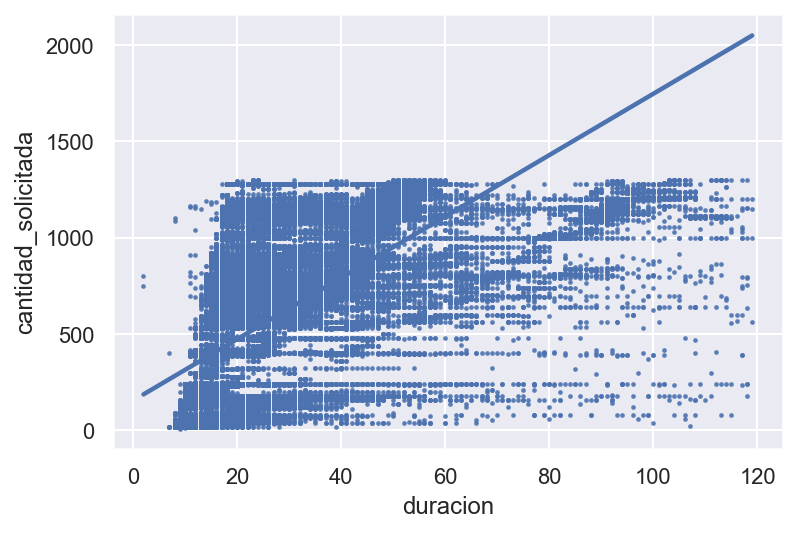
Creación de nuevas características:

* Hora. Hora a la que se produce la dosificación.
* DOW. Día de la semana en el que se produce la dosificación.
* Tmed. Temperatura media del día (OpenData AEMET).

Correlación entre variables





Cantidad solicitara – duración

Por silo y Materia

Por silo

Preprocesamiento:

Numéricas: Cantidad solicitada, peso inicial, tamaño mezcla y tmed.

Categóricas: Materia origen, Materia destino, silo, orden, hora, dow.

Train y test: 66% y 33%

Entrenamiento de modelos:

|  |  |  |  |
| --- | --- | --- | --- |
| Modelo | Error Train | Error Test | Hiperparametros |
| R. Lineal | 4.18 Seg. | 4.15 Seg. |  |
| Ridge | 4.18 Seg. | 4.15 Seg. | Alpha = 1 |
| Random Forest | 1.31 Seg. | 1.55 Seg. | Estimadores = 200  Max Atributos = ‘auto’  Max Profundidad = 20 |
| Random Forest 1 silo | 0.39 Seg. | 0.87 Seg. |
| Random Forest 1 silo 1 materia | 0.35 Seg. | 0.88 Seg. |
|  |  |  |  |

A B C

Paso a producción:

Conclusiones:

* Importancia de la fase de análisis.
* Importancia Generación de características.
* Random Forest funciona perfectamente con gran cantidad de datos y variables.
* En nuestro problema muy caracterizado por el diseño y la disposición del elemento mecánico.

Trabajos futuros:

* Modelo de clasificación de materia en silos en función de menor desviación en la dosificación.

GRACIAS POR SU ATENCIÓN