Trabajo Fin de Máster

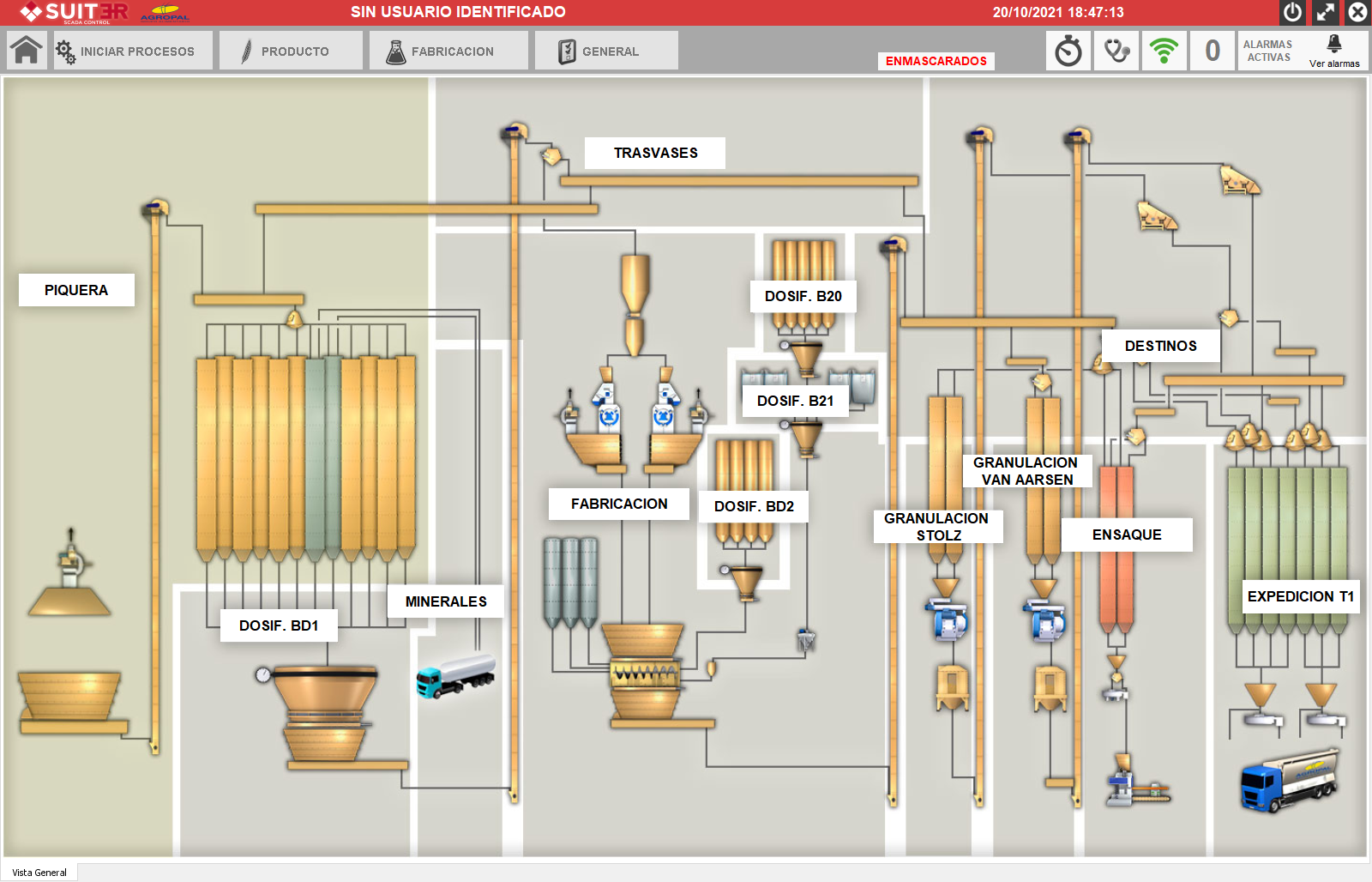
Máster en Ciencia de Datos e Ingeniería de Datos en la Nube

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

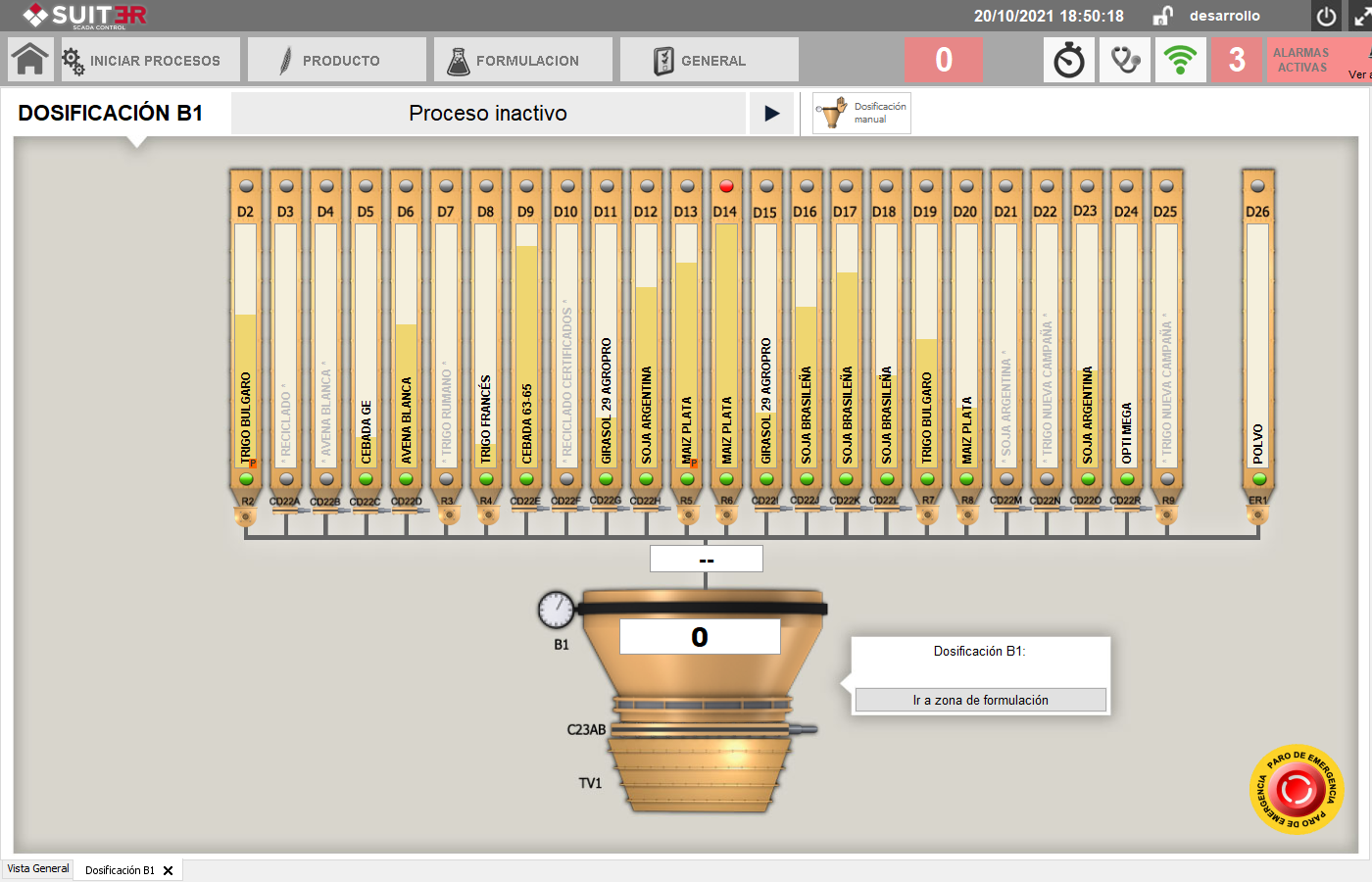
Auto: Jose Luis Casado Valero

Tutor: Luis de la Ossa Jiménez

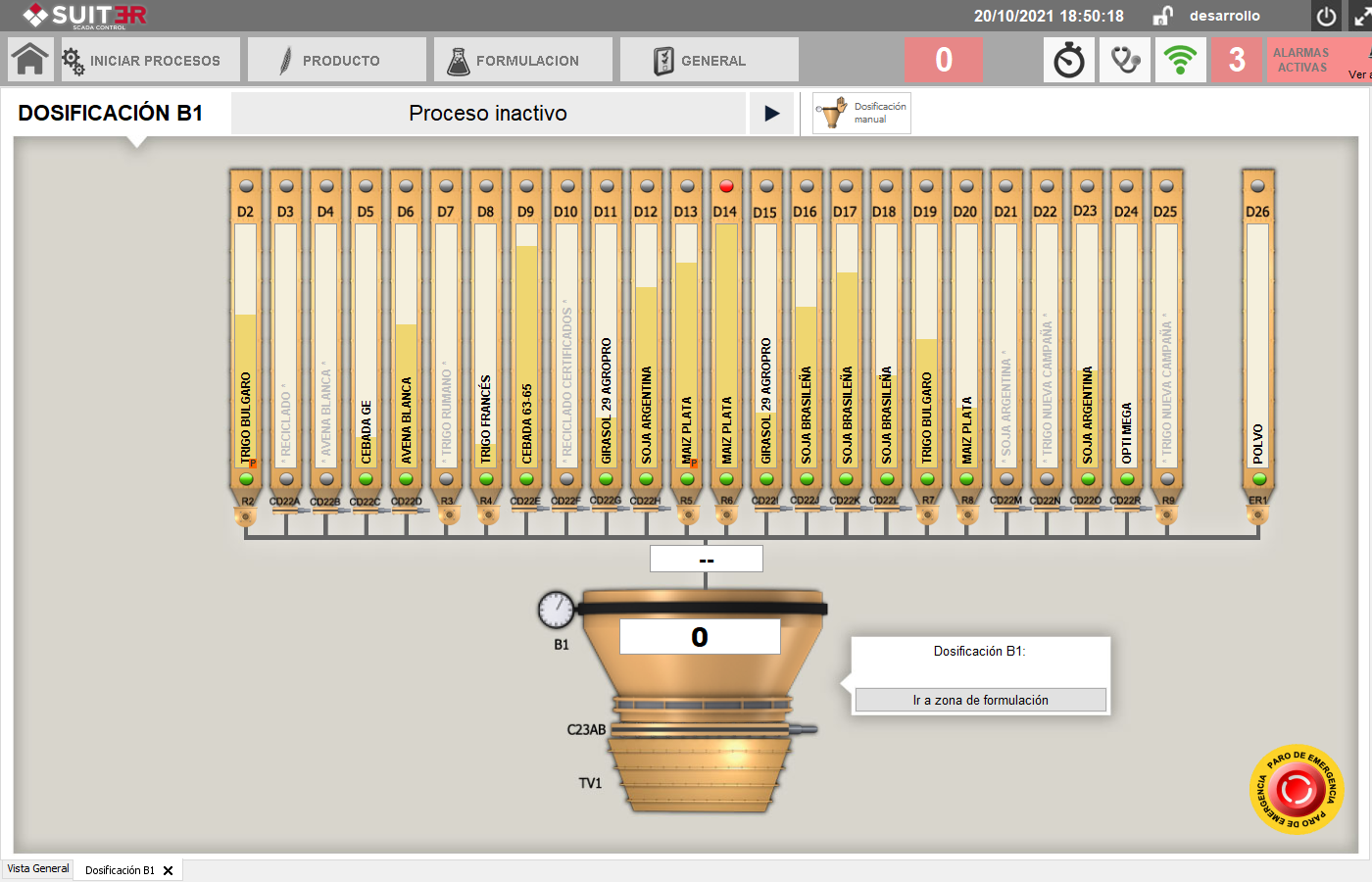
**Fabrica de piensos:**



**Dosificación en bascula:**



**Dosificación en bascula:**

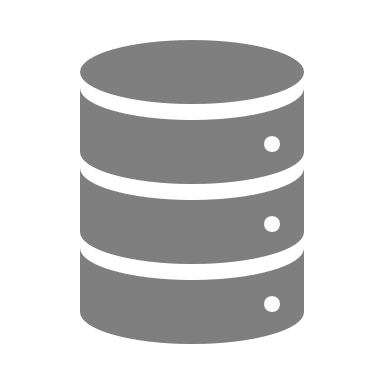


DEMO

|  |  |
| --- | --- |
| Nombre columna | Ejemplo |
| Fecha\_inicio | 2013-08-14 16:51:48 |
| Fecha\_fin | 2013-08-14 16:52:39 |
| Cantidad\_solicitada | 235.026 |
| Cantidad\_dosificada | 236.5 |
| Mezcla | 1 |
| Pesada | 0 |
| Peso\_inicial | 0.0 |
| Manual | 0 |
| Id\_lote\_destino | 232527 |
| Materia\_origen | 13 |
| Materia\_destino | 66 |
| Id\_silo | 90 |
| Tipo\_materia | Prima |
| Tipo\_destino | Premezcla |
| Densidad | 1.0 |
| Tam\_mezcla | 500 |
| Desviacion | 1.47 |

**Fuente de datos:**

* Registros de trazabilidad aplicación de automatización.
* 395.477 registros.
* No existen valores perdidos.



**Objetivo:**

* Esgrimir los datos en busca de información para entender las desviaciones.
* Prevenir las desviaciones en las dosificaciones.
* Conseguir un método complementario al cálculo de colas actual.

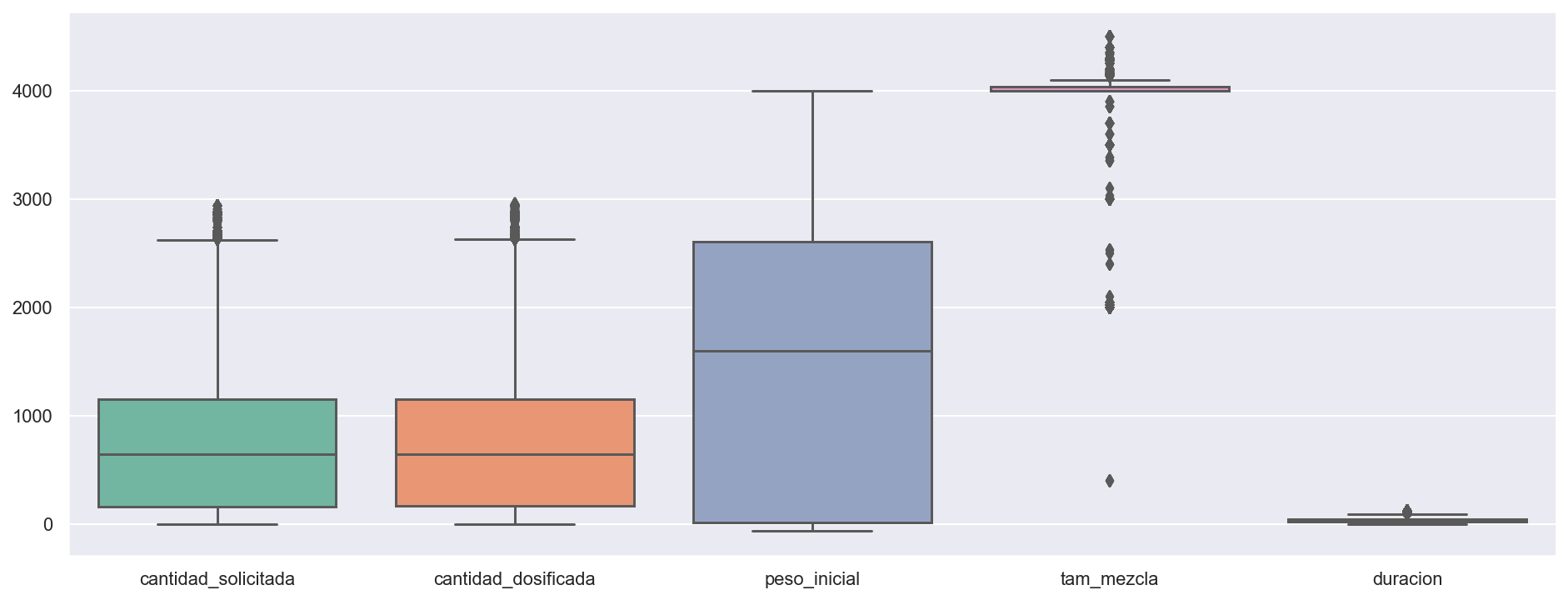
**Planteamientos:**

* Predicción del error. (Difícil de caracterizar)
* Predicción de duración.
  + Corrección de cola condicionada a la duración.

**Duración:**

* Variable a predecir.
* Segundo que está activo el elemento dosificador.
* Dataset. Segundos que transcurren desde fecha inicio a fecha fin.
* Ejemplo: 51 seg.

**Estadísticos variables numéricas:**



|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | 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.

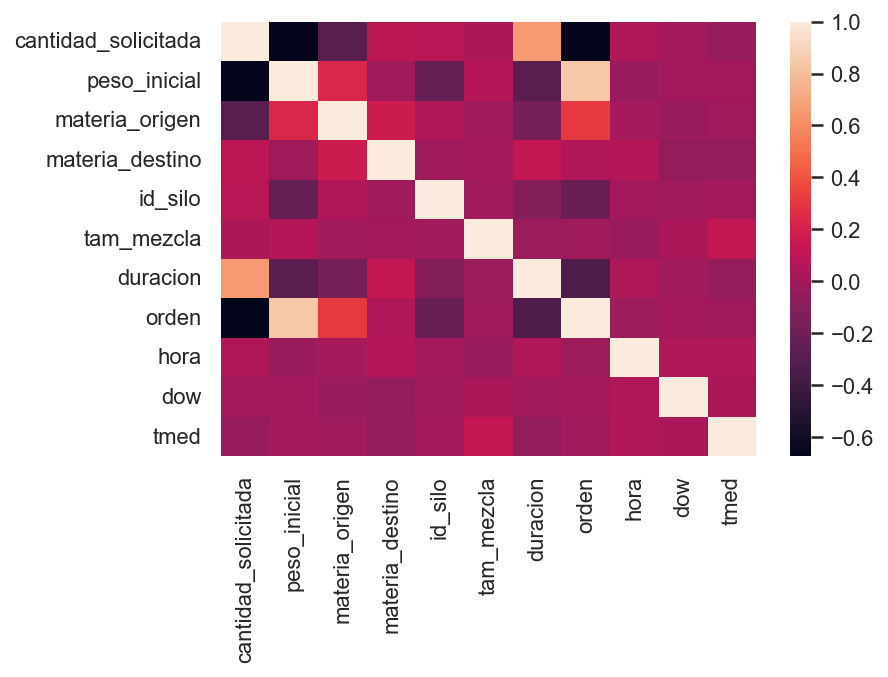
**Filtrado de columnas:**

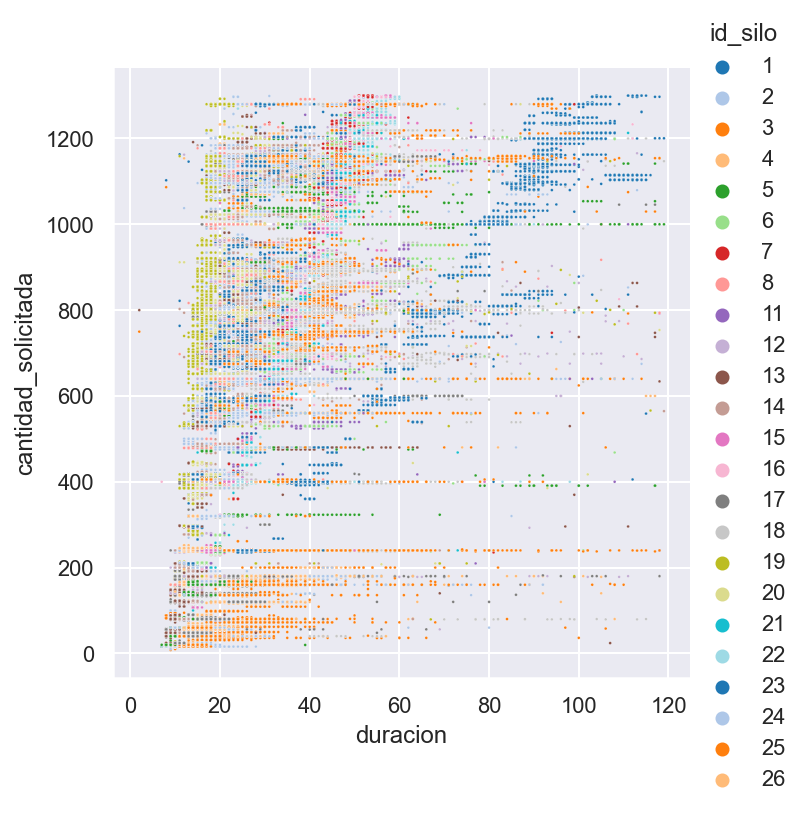
* Fecha\_fin.
* Cantidad\_dosificada.
* Mezcla.
* Pesada.
* Manual.
* Densidad.
* Tipo\_materia.
* Tipo\_destino.
* Desviación.

**Creación de nuevas características:**

* Orden. Orden de cada dosificación dentro de cada mezcla.
* 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:**

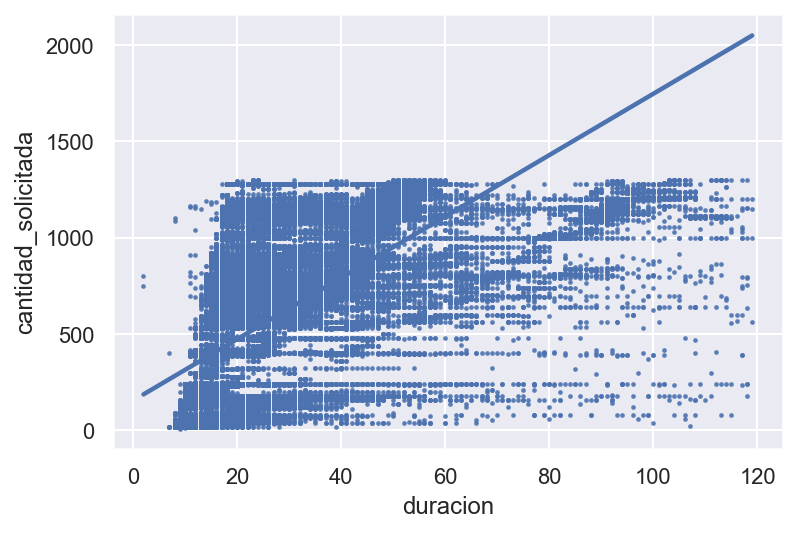
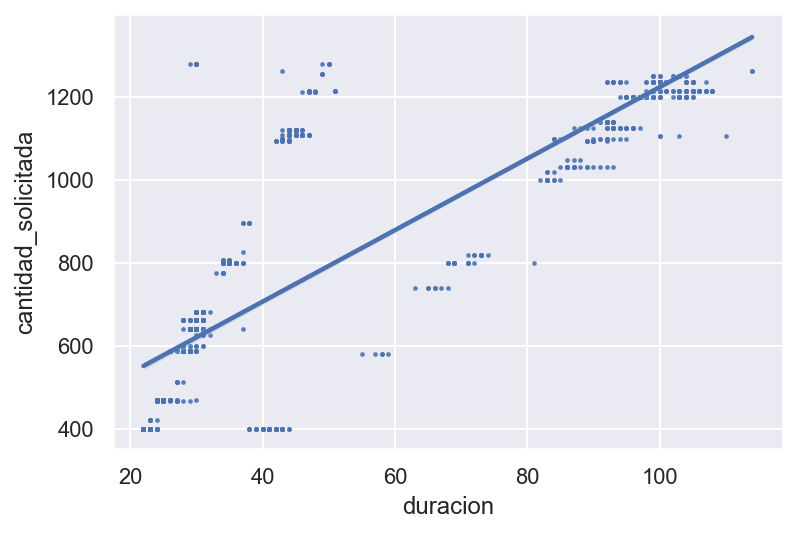




**Cantidad solicitada – duración:**

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:**

* Aleatorio. 66% y 33%
* Curiosidad: Sobreajuste al hacerlo cronológicamente.

**Entrenamiento de modelos:**

* Utilizamos le error absoluto medio.
  + Cuantificar y comparar modelos
  + Medida de valoración desde el punto de vista operativo.

|  |  |
| --- | --- |
| Modelo | C.V búsqueda de Hiperparametros |
| R. Lineal |  |
| Ridge | Alpha = {**1**, 0.1, 0.001, 0.0001} |
| Random Forest | Estimadores = {50, 100, **200**, 1000}  Max Atributos = **‘auto’**, 3, 5, 7, 10  Max Profundidad = None, 3, 5, 10, **20** |
| Random Forest 1 silo |
| Random Forest 1 silo 1 materia |
|  |  |

**Resultados:**

|  |  |  |  |
| --- | --- | --- | --- |
| 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

**Implantación en 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, la dosificación está muy caracterizada 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