

Coderhouse – Data Scientist

Segunda entrega proyecto final

Emilse Bover

Abstract

En presente trabajo se analizaron los resultados productivos de una empresa productora de pollos parrilleros para detectar las variables que impactan sobre los resultados y realizar recomendaciones a partir de ello. En segunda instancia se construyó un modelo de predicción de resultados productivos en función de los datos que se suministren en el futuro.

Objetivo

El objetivo del presente trabajo es detectar el impacto de los factores que afectan al crecimiento de los pollos parrilleros para tomar decisiones que mejoren los resultados productivos.

Contexto comercial

El retorno económico de la empresa productora de pollos parrilleros depende de la obtención de mejores resultados productivos. A mayor ganancia de peso diaria, por ejemplo, el consumo total de alimento será menor disminuyendo el costo. Otro impacto positivo del aumento de la ganancia de peso diaria es que se reduce la edad a faena de los animales ya que se alcanza el peso deseado en menor cantidad de días. Esta reducción de días permite: liberación de superficie de producción (metros cuadrados de galpón, rotación), menor propensión a sufrir enfermedades y accidentes (por ejemplo cortes de luz). También el dueño de la granja tiene menores costos de luz y gas y por lo tanto mayores ganancias finales también.

Contexto analítico

```
# Se presentan las librerías utilizadas:  
import pandas as pd  
import xlswriter  
import matplotlib as mpl  
import matplotlib.pyplot as plt  
import seaborn as sns  
import scipy as sp
```

In []:

```
import numpy as np
import datetime
import pingouin as pg
import statsmodels.api as sm
from scipy import stats
import requests
import json
```

```
!pip install nbconvert
! pip install -U scikit-learn
```

```
! pip install pyppeteer
```

```
import plotly.express as px
```

In []:

```
pip install https://github.com/pandas-profiling/pandas-
profiling/archive/master.zip
```

In []:

```
import ydata_profiling
```

Data acquisition

La empresa ha provisto un archivo excel con información de resultados y se utilizaron datos climáticos históricos obtenidos a través de una API de Open Meteo.

In []:

```
#Ingresos semanales compilado registra resultados finales de la
producción en la pestaña BASE. Contiene además datos de genética,
nutrición y sanidad.
df = pd.read_excel(r"C:\Users\ebover\OneDrive - FRIGORIFICO DE AVES
SOYCHU S.A.I.C.F.I.A\ingresos semanales
compilado.xlsx",sheet_name='BASE', skiprows=1 , usecols=range(1,67))
```

In []:

```
# Se hizo una solicitud HTTP a la API de Open-Meteo
response = requests.get('https://archive-api.open-
meteo.com/v1/archive?latitude=-32.48&longitude=-58.23&start_date=2020-
01-01&end_date=2022-12-
31&daily=temperature_2m_max,temperature_2m_min,precipitation_sum&timez
one=America%2FSao_Paulo')
```

```
# Se convirtieron los datos JSON en un diccionario de Python
data = response.json()['daily']
```

```
# Se convirtió el diccionario en un DataFrame de pandas
dfc = pd.DataFrame(data)
```

Data wrangling

Primero se eliminaron valores erróneos de la base de datos que se conocían antes de iniciar el análisis:

Base de datos productivos

La base de datos contiene resultados productivos de 15.571 crianzas desde el año 2015 hasta la primera quincena del mes de abril de 2023.

Los valores en blanco hallados pertenecen a una columna que no será utilizada en este análisis

La base no posee valores de crianzas duplicados.

Base de datos climáticos de Concepción del Uruguay, Entre Ríos

La base no posee valores en blanco

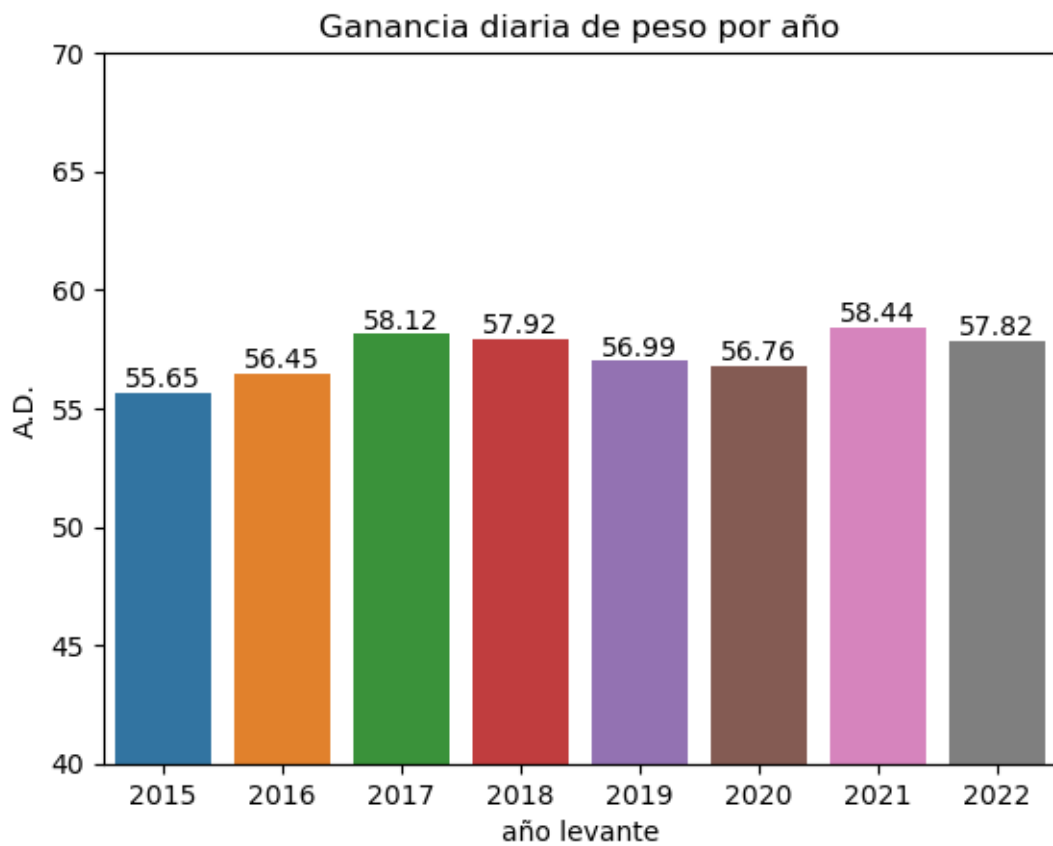
La base no posee fechas duplicadas

De la base de datos provista por la empresa se seleccionarán las siguientes columnas para el análisis:

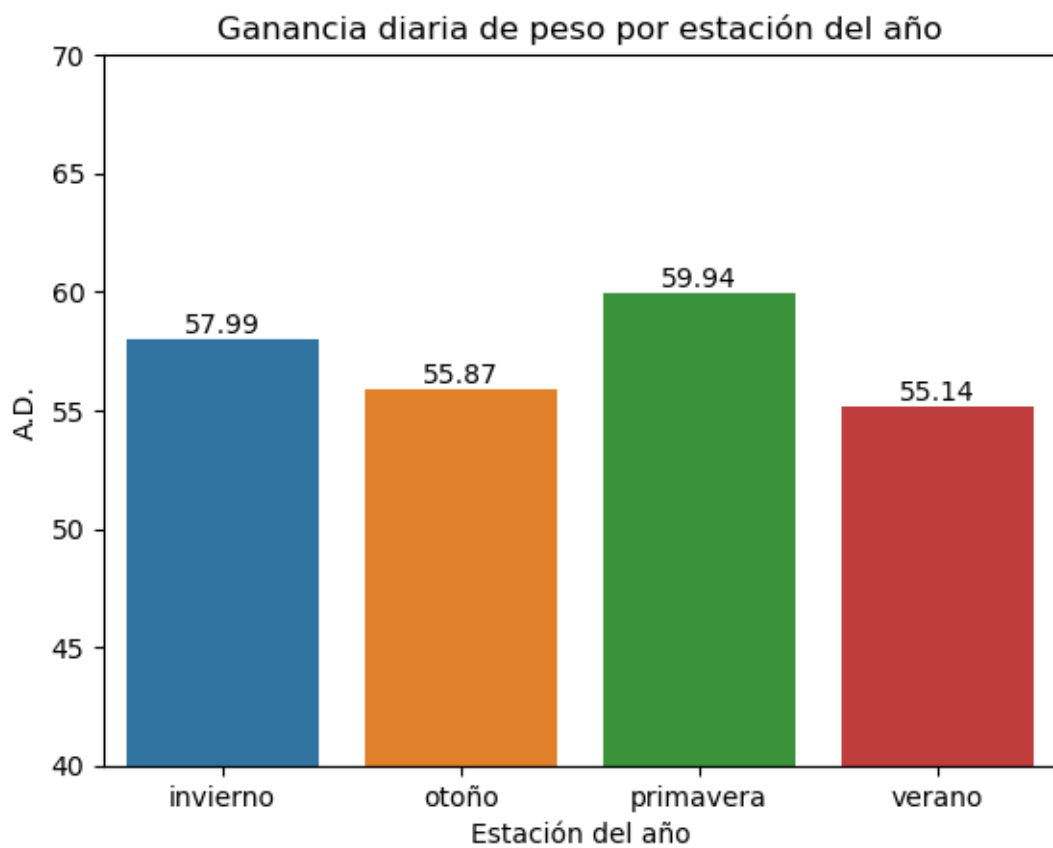
Nombre	Descripción
granja	Unidad productiva compuesta por galpones
Primer BB	Fecha inicio crianza de aves
A Faena	Cantidad de aves al final de la crianza
% Mortan.	Mortalidad animales
P. Prom.	Peso promedio a faena
Ulto. Levan.	Fecha fin de crianza de aves
Edad	Edad al final de la crianza
A.D.	Ganancia media diaria de peso
año levante	año de fin de crianza
mes levante	mes de fin de crianza
Estación del año	Momento del año en que ocurrió el fin de la crianza
Consumo medio diario (g)	Consumo de alimento por ave
Destino	Destino de venta: mercado interno o exportación
Densidad	Cantidad de aves por metro cuadrado de galpón (de cada granja)
kg pollo/m2	Cantidad de kilos logrados por unidad de superficie de granja
Ventilación forzada	Tipo de sistema de ventilación de granja
Ambiente controlado	Subtipo de sistema de ventilación
localidad	Localidad geográfica de las granjas
% desvío consumo std	Desvío del consumo de alimento respecto del estándar
Índice	Número de crianza (único)
zona climática	Zona de producción según clima imperante
Pes conv 2,7	Indicador de eficiencia productivo
zona prod	Zona de producción de la granja

Análisis exploratorio de datos

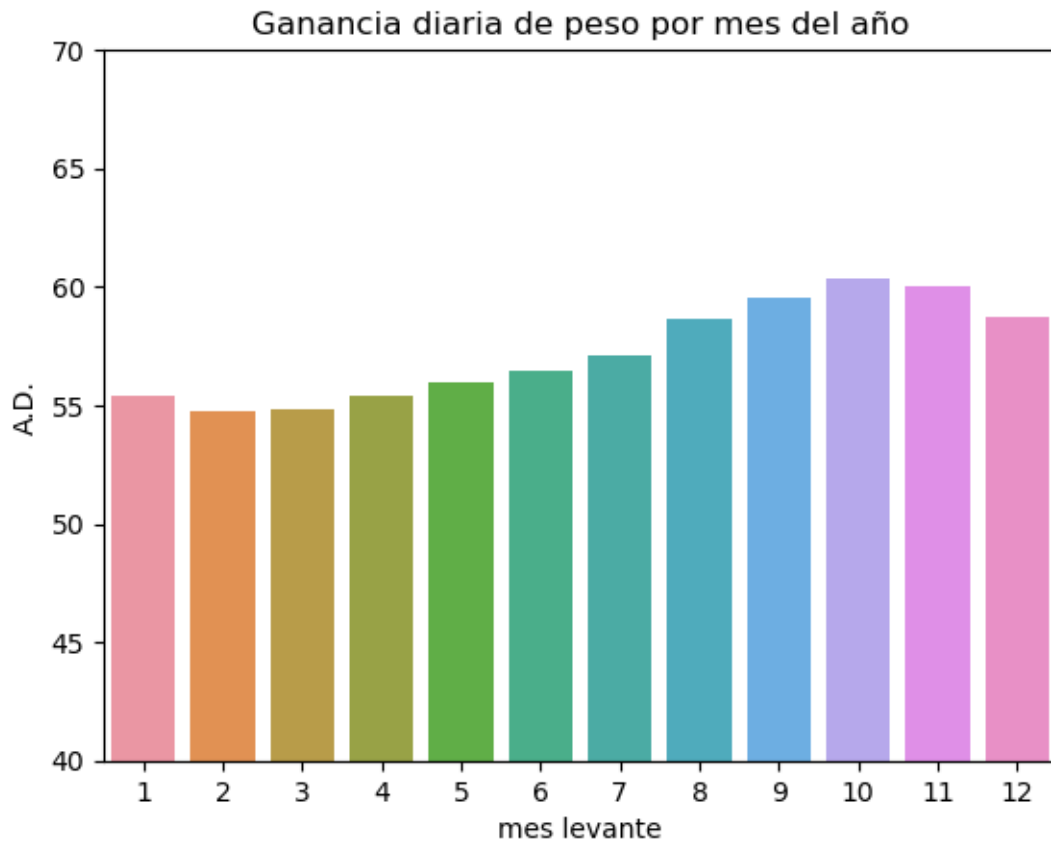
La **ganancia media diaria de peso** es una medida que permite comparar resultados entre granjas y además permite ver la evolución de los resultados globales de la empresa a lo largo del tiempo. A mayor valor, mejor será el resultado. El valor medio es de **57.18** gramos de peso ganados promedio por día.



Se observa que las ganancias de peso de los años 2017 y 2021 fueron superiores al resto de los años analizados. El año 2023 fue excluido por estar incompleto.



Se observa que la ganancia de peso es variable según la estación del año en la que se desarrolle la crianza, las mayores ganancias de peso ocurren en primavera y las peores en verano dando indicios de que el ambiente puede estar afectando el resultado productivo.



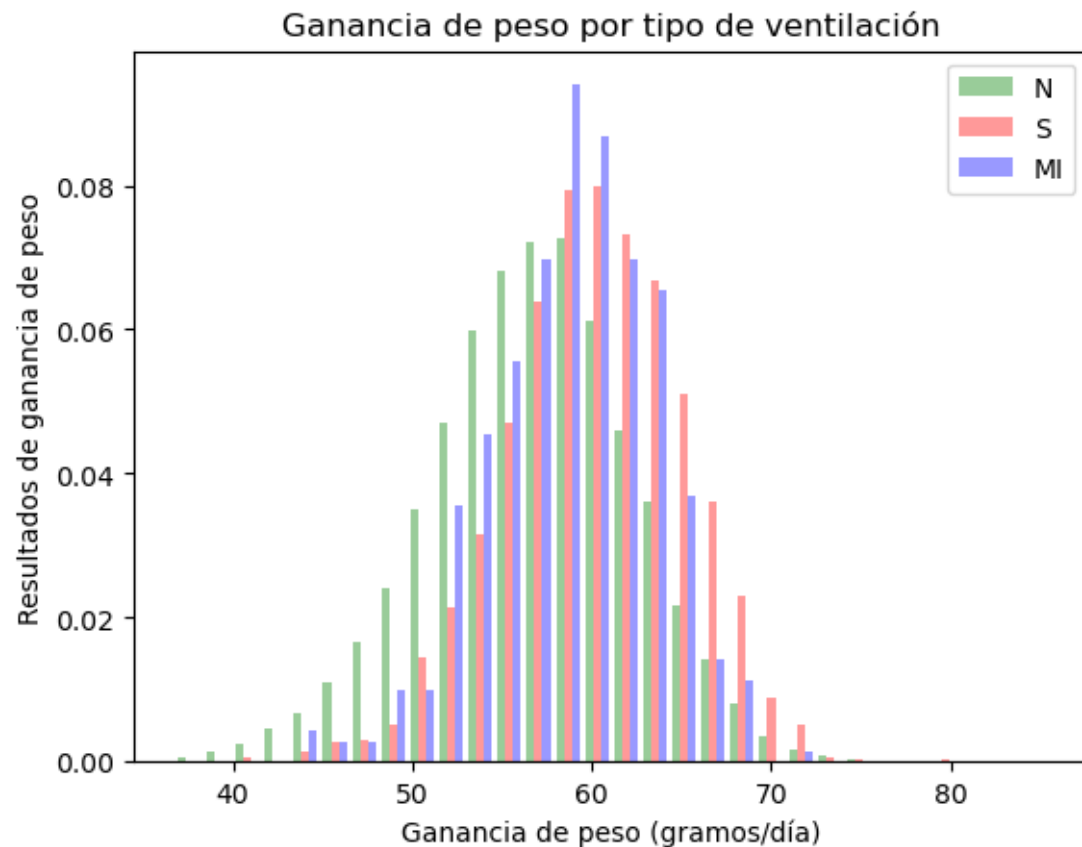
Se observan también diferencias en los valores de ganancia media diaria de peso en los meses del año. Debido a ello a continuación se realizará un análisis según los tipos de ventilación que poseen las granjas.

Para ello se han clasificado las granjas en tres categorías:

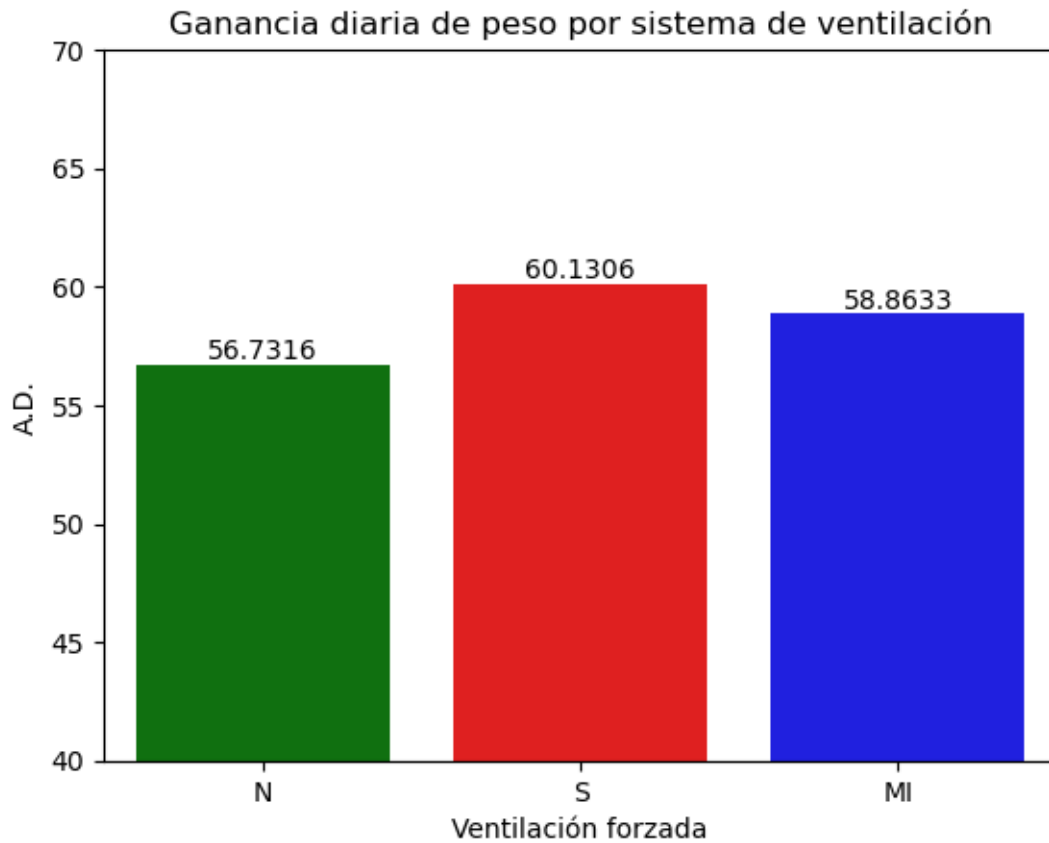
S: Ventilación forzada: La ventilación del galpón se realiza a través de la entrada del aire por inlets o aberturas en la zona superior de los galpones y su posterior expulsión por extractores.

N: Granjas con sistema de ventilación convencional con encendido manual de ventiladores y/o apertura y cierre de las cortinas laterales del galpón.

MI: Mixtas: Granjas que tienen ambos tipos de galpones.



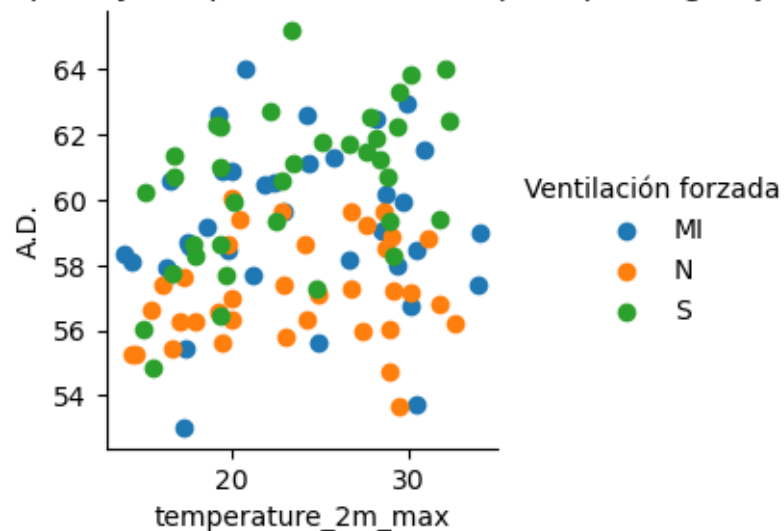
En el gráfico se observa que los valores de ganancia de peso de los tres sistemas tienen una distribución normal y que el sistema convencional logra menores ganancias de peso que los sistemas de ventilación forzada y mixto. La distribución de los datos de las granjas convencionales se encuentra desplazada hacia la izquierda, es decir hacia valores menores de ganancia. Su promedio se ubica alrededor de los 56 gramos/días mientras que el promedio de las granjas de ventilación forzada está alrededor de los 60 gramos/día.



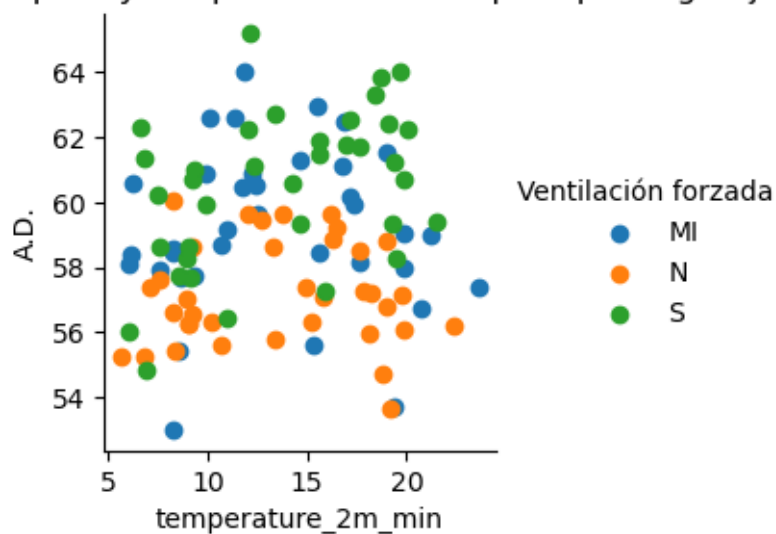
Habría indicios de que el sistema de ventilación tiene impacto sobre la ganancia de peso de los animales.

#Genero un dataframe con los datos filtrados para la localidad de Concepción del Uruguay y fechas para los que se tienen valores de temperaturas

Ganancia de peso y temperatura máxima por tipo de granja

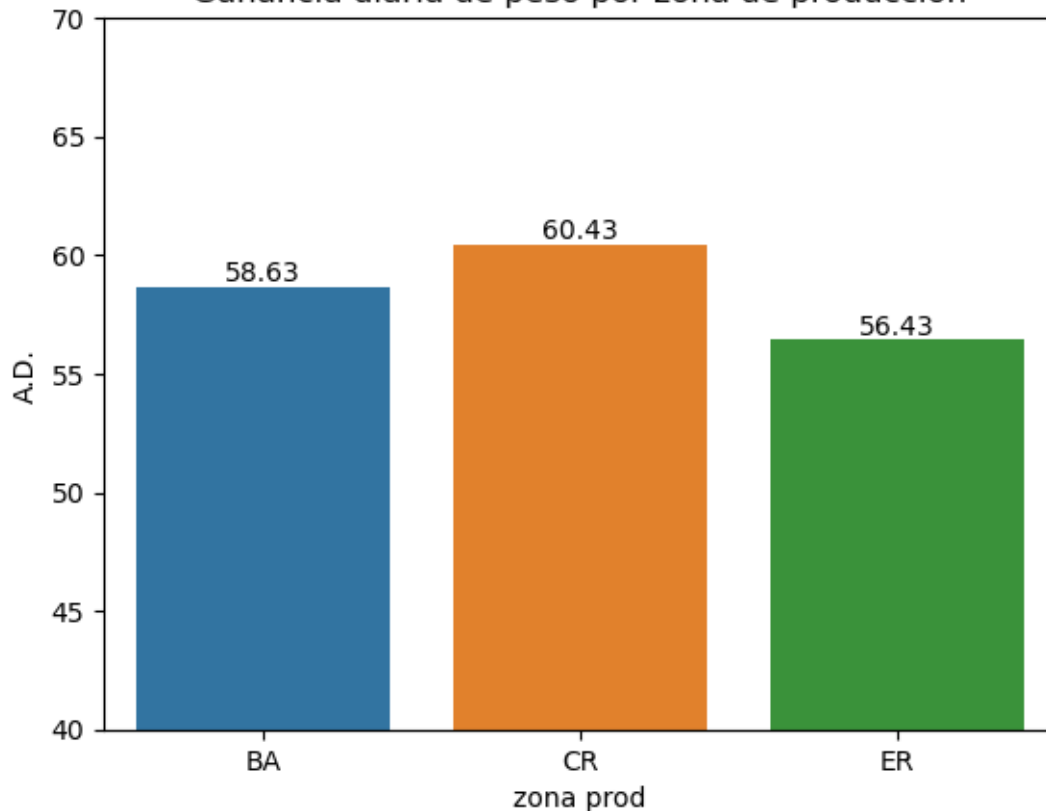


Ganancia de peso y temperatura mínima por tipo de granja

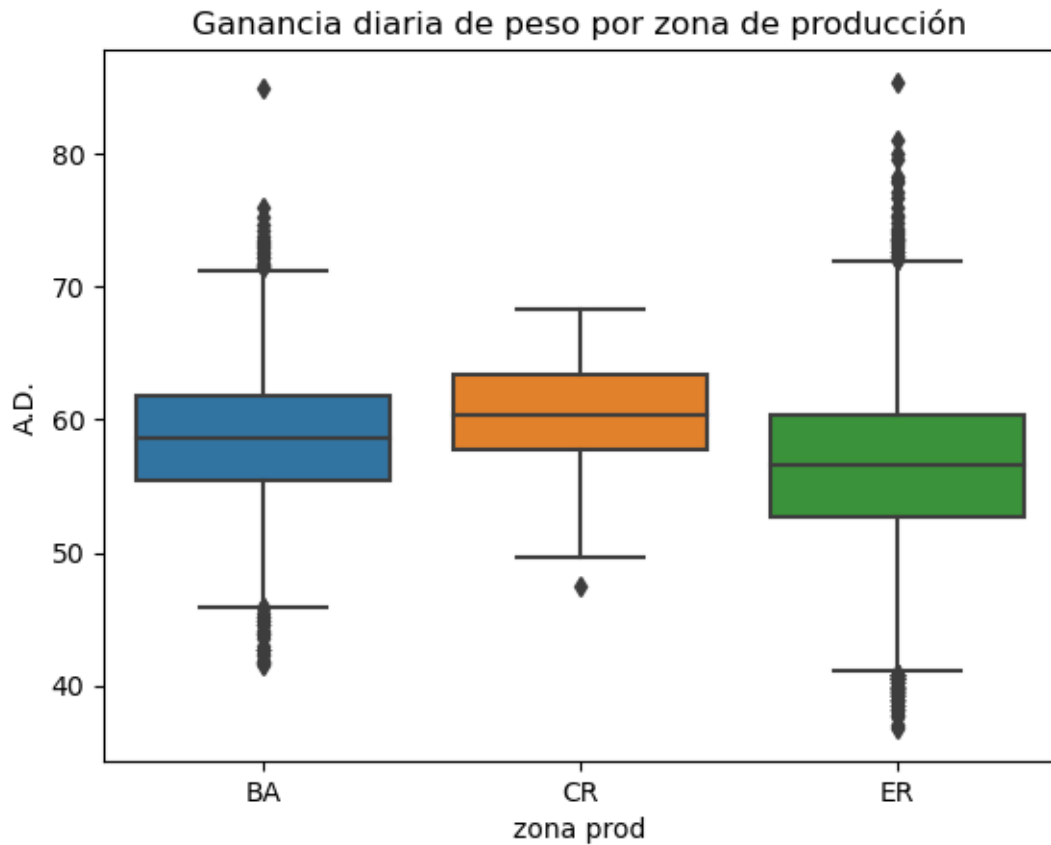


Los gráficos de correlación entre la ganancia de peso y las temperaturas máximas y mínimas diarias no muestran ninguna relación entre estas variables y tampoco en relación al sistema de ventilación.

Ganancia diaria de peso por zona de producción



A nivel de promedios, hay diferencias entre las zonas del país donde se encuentran ubicadas las granjas.



Los mejores resultados se encuentran en la zona de Crespo (CR), no solo porque tiene mayores valores de ganancia de peso sino también porque los resultados tienen menor dispersión y prácticamente no posee resultados atípicos.

Machine learning aproximación

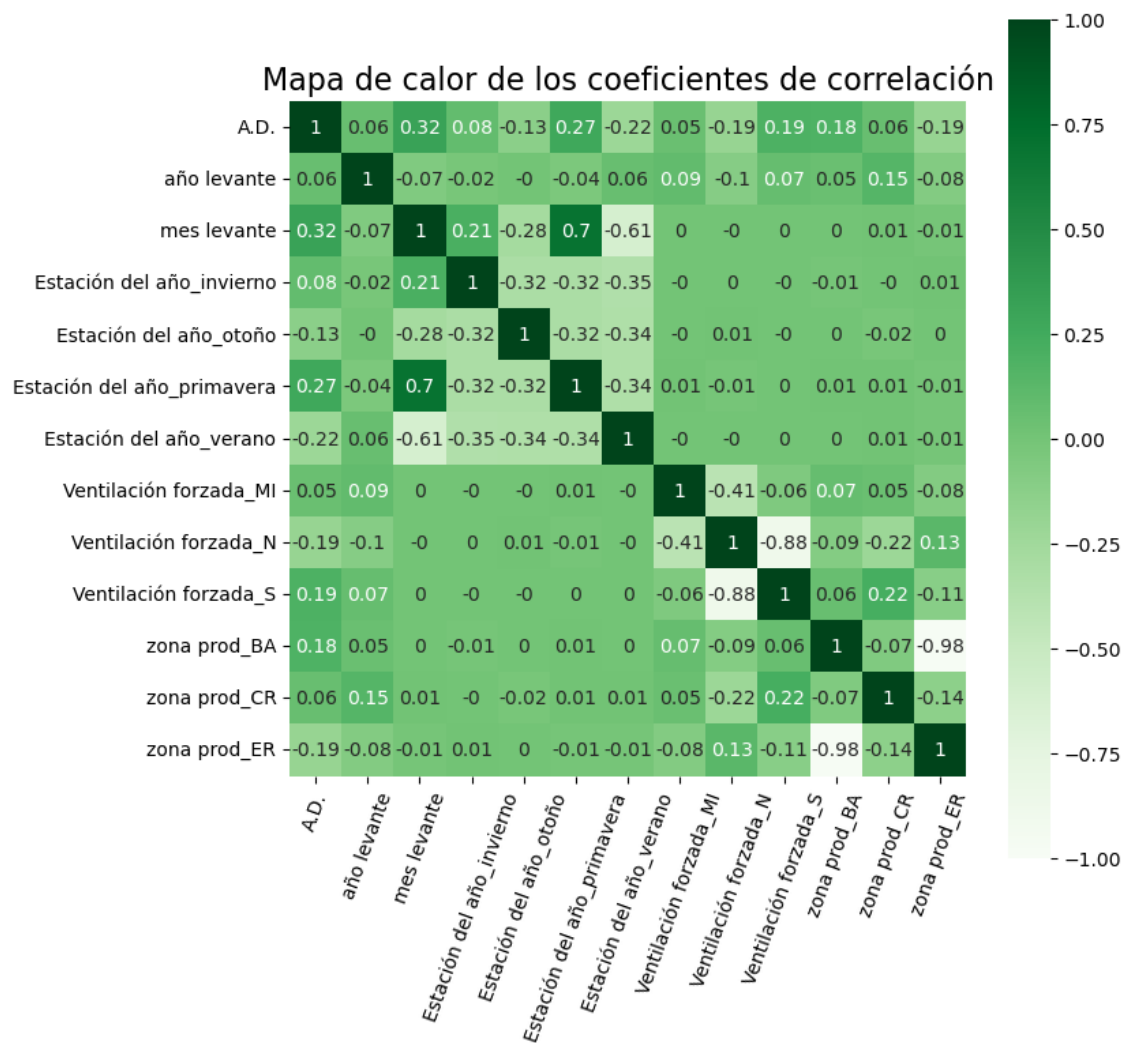
Modelo SVR

Error cuadrático medio (MSE): 25.104881337804418

Coefficiente de determinación (R^2): 0.24259967583607478

El modelo aplicado tiene un error cuadrático medio de 25.1 con un coeficiente de determinación del 24.25%.

Correlaciones



Del análisis del mapa de calor se desprende que la correlación entre la ganancia diaria de peso y las variables analizadas es baja en general, siendo los mayores valores un 32% para el mes de levante, un 27% de correlación positiva para la estación del año primavera y un 22% de correlación negativa para la estación verano.