Informe técnico para la empresa Délicieux

Estimado equipo de Délicieux, es un placer anunciarle que ya no tendrá que preocuparse por si los distintos tipos de hongos que va a usar en sus platos son tóxicos o no, gracias al dataset que nos proporcionó hemos detectado patrones interesantes y el próximo paso es crear una plataforma móvil para que clasifique según los datos de color, textura y olor.
A continuación desglosare lo más importante:

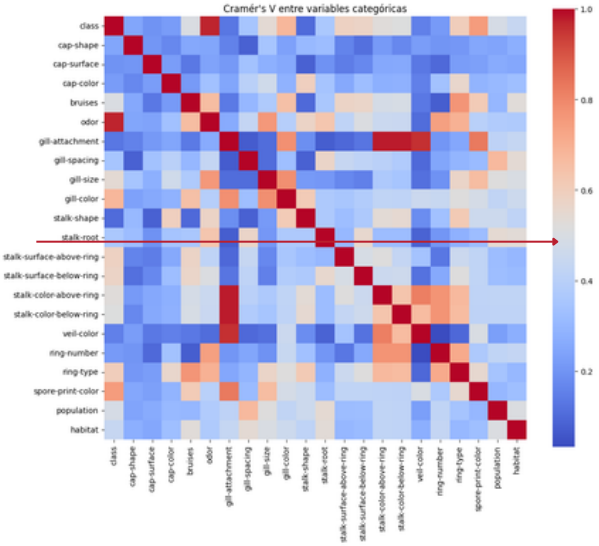
Análisis exploratorio de los datos

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 8124 entries, 0 to 8123
Data columns (total 23 columns):
 #   Column                    Non-Null Count  Dtype 
---  ------                    --------------  ----- 
 0   class                     8124 non-null   object
 1   cap-shape                 8124 non-null   object
 2   cap-surface               8124 non-null   object
 3   cap-color                 8124 non-null   object
 4   bruises                   8124 non-null   object
 5   odor                      8124 non-null   object
 6   gill-attachment           8124 non-null   object
 7   gill-spacing              8124 non-null   object
 8   gill-size                 8124 non-null   object
 9   gill-color                8124 non-null   object
 10  stalk-shape               8124 non-null   object
 11  stalk-root                8124 non-null   object
 12  stalk-surface-above-ring  8124 non-null   object
 13  stalk-surface-below-ring  8124 non-null   object
 14  stalk-color-above-ring    8124 non-null   object
 15  stalk-color-below-ring    8124 non-null   object
 16  veil-type                 8124 non-null   object
 17  veil-color                8124 non-null   object
 18  ring-number               8124 non-null   object
 19  ring-type                 8124 non-null   object
 20  spore-print-color         8124 non-null   object
 21  population                8124 non-null   object
 22  habitat                   8124 non-null   object
dtypes: object(23)
memory usage: 1.4+ MB
None
```

El dataset proporcionado constaba de 8124 datos divididos en 23 columnas todas categóricas

En cuanto a nulos aparentemente no había, sin embargo cuando vemos valores extraños nos encontramos con ? en la variable stalk-root con 2480 datos

```
Valores '?' por columna:
 class                       0
cap-shape                    0
cap-surface                  0
cap-color                    0
bruises                      0
odor                         0
gill-attachment              0
gill-spacing                 0
gill-size                    0
gill-color                   0
stalk-shape                  0
stalk-root                2480
stalk-surface-above-ring     0
stalk-surface-below-ring     0
stalk-color-above-ring       0
stalk-color-below-ring       0
veil-type                    0
veil-color                   0
ring-number                  0
ring-type                    0
spore-print-color            0
population                   0
habitat                      0
dtype: int64
Valores '?' en 'stalk-root' reemplazados por la moda: b
```

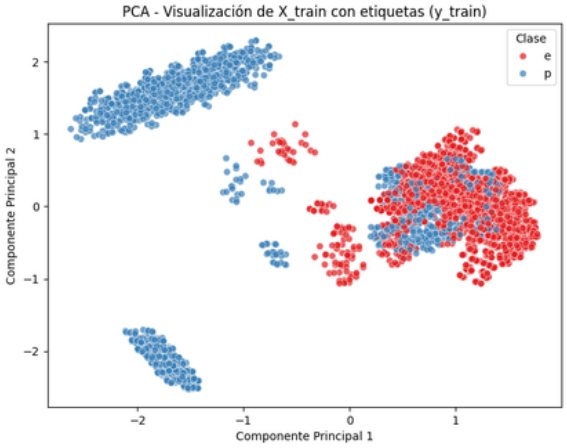


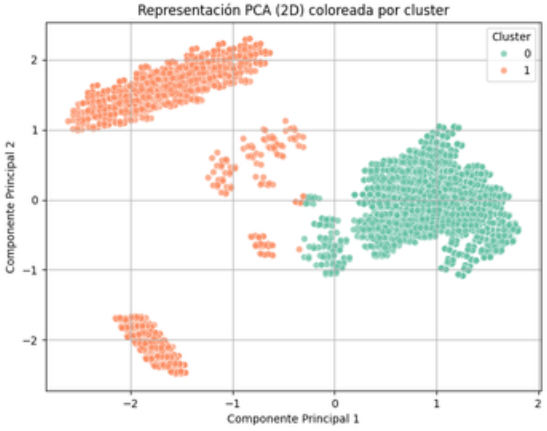
Como se puede observar en la V cramér esta variable no posee tanta correlación. Con lo cual no afecta al entrenamiento del modelo.

En cuanto al análisis de los datos proporcionados esto es lo más relevante. Cabe destacar que la Variable objetivo es class ya contiene en binario si es comestible o venenoso

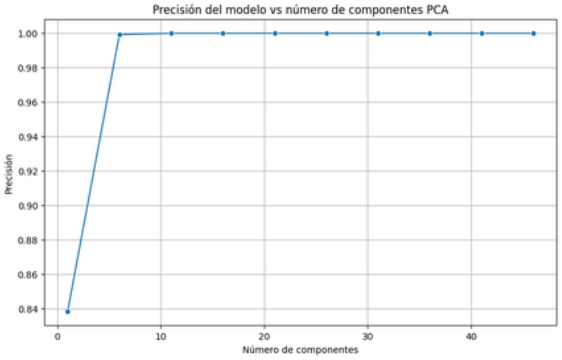
Algoritmo de entrenamiento

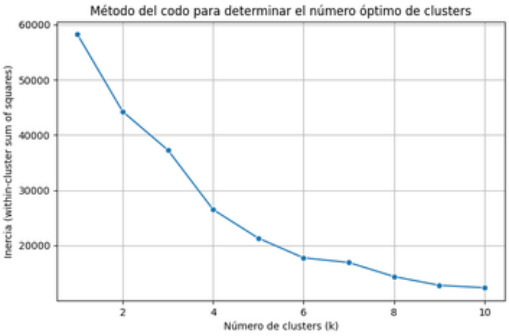
Se realizó un PCA para reducir el ruido, acelerar el entrenamiento del modelo y sobretodo evitar la multicolinealidad y por lo que se observa el antes y el después de haber sido entrenado.



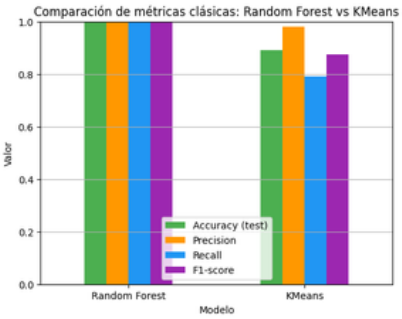


Para conseguir esto primero visualizamos el score donde mejor entrenamiento tiene el modelo y vemos que es a partir del 10, luego con el método del codo vemos que entre 2-4 clusters funcionaria muy bien así que elegimos K2.





Para el modelo final elegimos Random Forest ya que es el que mejor score nos da.



	Modelo	Accuracy (test)	Precision	Recall	F1-score
0	Random Forest	1.000000	1.000000	1.000000	1.000000
1	KMeans	0.892418	0.981634	0.791624	0.876449

Quedamos a su disposición para proceder a la creación de la plataforma móvil.
Un saludo de parte del equipo de ShroomBuster