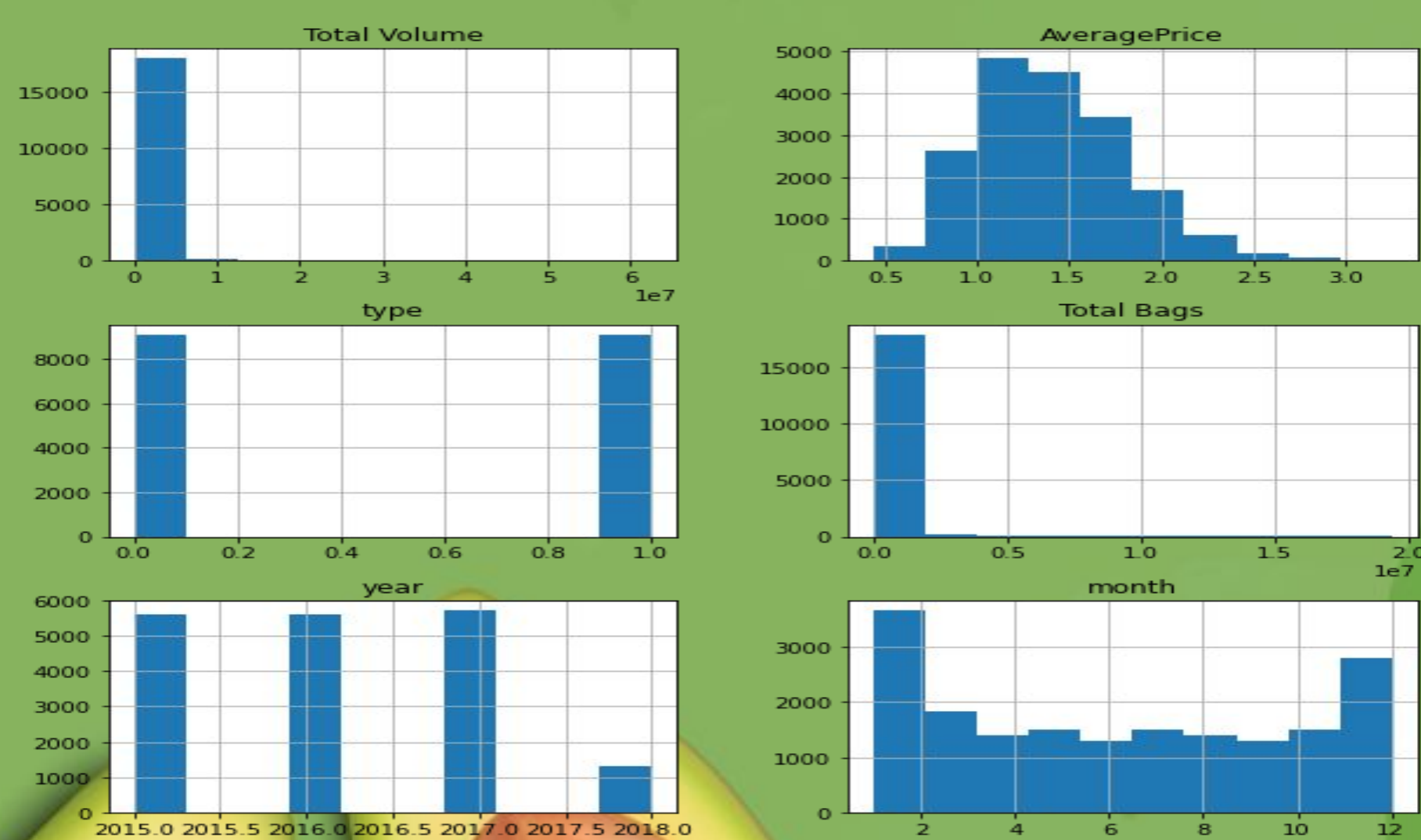
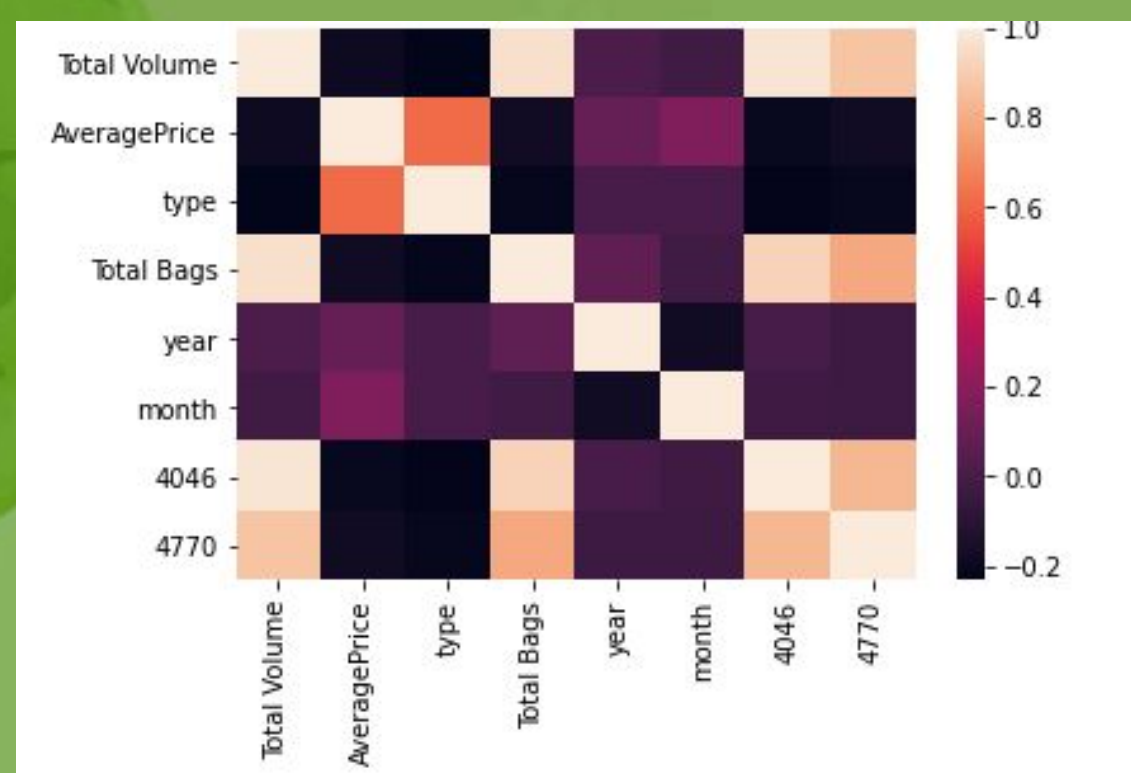


AGUACATE

PRODUCTO INSIGNIA PARA SER POTENCIA

Proyecto final - Introducción a los Sistemas Inteligentes
2020-II

EXPLORACIÓN DE DATOS



- Al revisar varias de las métricas como **precision** y **recall** altas nos da un **f1** alto, y esto quiere decir que es un excelente clasificador.
- La exploración aleatoria arroja unos estimadores bastante efectivos dado que con el mejor estimador y observando la matriz de confusión arroja un solo falso negativo y un falso positivo, como se ve en las gráficas de desempeño

INTRODUCCIÓN:

A continuación se realiza una revisión de algunos datos sobre el aguacate y se pretenden exponer sus beneficios económicos, entre otras razones, para elegirlo como un producto con un gran potencial de exportación a nivel Colombia.

METODOLOGÍA Y ANÁLISIS:

- Recolección de datos provenientes de la plataforma Kaggle.
- Previsualización de los datos de origen
- Conversión de variables categóricas en numéricas y descarte de variables sin interés
- Relaciones entre variables
- Extracción de la clase
- Implementación *Random Forest*
- Toma de tiempos
- Creación de matriz de confusión
- Cálculo de porcentaje de acierto

MODELADO:

1. Se implementa el modelo **Random Forest**, dado que este se ajusta bien a la cantidad de datos y es por excelencia uno de los algoritmos más utilizados para un problema de clasificación como el propuesto.
2. Se usa **RandomizedSearchCV** para la optimización aleatoria de parámetros, porque empíricamente es más eficiente que **Grid Search**; para los datos utilizamos 30% para pruebas y 70% para entrenamiento de manera estratificada.
3. Se utiliza **cross validation** con 5 pliegues y 20 iteraciones y el
4. tiempo que tomó el entrenamiento es alrededor de 25 minutos.

```
random_search.fit(X_train, y_train)
print("RandomizedSearchCV tomó {} segundos usando {} configuraciones".format(time() - start,
                                                                              n_iter_search))

Fitting 5 folds for each of 20 candidates, totalling 100 fits
[Parallel(n_jobs=-1)]: Using backend LokyBackend with 2 concurrent workers.
[Parallel(n_jobs=-1)]: Done 37 tasks      | elapsed: 7.0min
[Parallel(n_jobs=-1)]: Done 100 out of 100 | elapsed: 18.0min finished
RandomizedSearchCV tomó 1089.6048738956451 segundos usando 20 configuraciones
```

CONCLUSIONES:

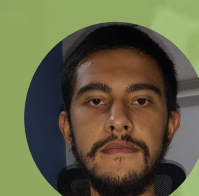
- El modelo es bastante acertado y tiene el potencial de permitir análisis más detallados para que el gobierno se enfoque en la producción de un tipo de aguacate para exportar.
- Luego de estudiar los datos anteriores se evidencia que el aguacate sí es un producto capaz de volver potencia económica a un país como Colombia que lo conoce y maneja de primera mano.

Sergio Andrés González Martínez
andgonzalezmar@unal.edu.co

Sebastian Guerrero Salinas
sebguerrerosal@unal.edu.co

Rafael Antonio Salgado López
rasalgadol@unal.edu.co

Profesor Fabio Augusto González
fagonzalezo@unal.edu.co



UNIVERSIDAD
NACIONAL
DE COLOMBIA

100%
ORGANIC