Questão 1

In [28]:
```
from sklearn import datasets

data = datasets.load_diabetes()

X = data.data
y = data.target
```

In [29]:
```
from sklearn.model_selection import KFold

kf = KFold(n_splits=5)


for train_index, test_index in kf.split(X):
     X_train, X_test = X[train_index], X[test_index]
     y_train, y_test = y[train_index], y[test_index]
```

In [30]:
```
from sklearn.linear_model import LinearRegression
from sklearn.ensemble import RandomForestRegressor

from sklearn.metrics import mean_absolute_error, mean_squared_error, r2_score
```

In [31]:
```
reg = LinearRegression().fit(X_train, y_train)
y_pred= reg.predict(X_test)
```

In [32]:
```
reg.coef_
```

Out[32]:
```
array([  -3.74049412, -247.75926377,  520.40028736,  298.50772418,
       -637.82400819,  342.20759593,   35.39841929,  168.0798633 ,
        678.59201386,   87.56663187])
```

In [33]:
```
mae = mean_absolute_error(y_test, y_pred)
mae
```

Out[33]:
```
42.387107598312724
```

In [34]:
```
mse = mean_squared_error(y_test, y_pred)
mse
```

Out[34]:
```
2910.21268776043
```

In [35]:
```
r2 = r2_score(y_test, y_pred)
r2
```

Out[35]:
```
0.5502483366517519
```

In [36]:
```
rfr = RandomForestRegressor().fit(X_train, y_train)
y_pred= rfr.predict(X_test)
```

In [37]:
```
rfr.feature_importances_
```

Out[37]:
```
array([0.06506235, 0.01113841, 0.25387244, 0.10342889, 0.04610076,
       0.05156612, 0.05646213, 0.01955518, 0.32647518, 0.06633856])
```

In [38]:
```
mae = mean_absolute_error(y_test, y_pred)
mae
```

Out[38]:
```
50.79613636363636
```

In [39]:
```
mse = mean_squared_error(y_test, y_pred)
mse
```

Out[39]:
```
3823.0113477272735
```

In [40]:
```
r2 = r2_score(y_test, y_pred)
r2
```

Out[40]:
```
0.4091821124033559
```

Questão 2

In [41]:
```
import pandas as pd

df = pd.read_csv('diabetes.csv')
df.head()
```

Out[41]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	BMI	DiabetesPedigreeFunction	Age	Outcome
0	6	148	72	35	0	33.6	0.627	50	1
1	1	85	66	29	0	26.6	0.351	31	0
2	8	183	64	0	0	23.3	0.672	32	1
3	1	89	66	23	94	28.1	0.167	21	0
4	0	137	40	35	168	43.1	2.288	33	1

In [42]:
```
kf = KFold(n_splits=5, shuffle=True, random_state=42)

X = df.drop(columns=['Outcome']).values
y = df['Outcome'].values

for train_index, test_index in kf.split(X):
     X_train, X_test = X[train_index], X[test_index]
     y_train, y_test = y[train_index], y[test_index]
```

In [43]:
```
reg = LinearRegression().fit(X_train, y_train)
y_pred= reg.predict(X_test)
```

In [44]:
```
reg.coef_
```

Out[44]:
```
array([ 0.0209007 ,  0.00620197, -0.00195603,  0.00075591, -0.00026521,
        0.01126912,  0.14901865,  0.00225094])
```

In [45]:
```
mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

mae, mse, r2
```

Out[45]: (0.34289746131408305, 0.16607377176673602, 0.29444266365017746)

In [46]:
```
rfr = RandomForestRegressor().fit(X_train, y_train)
y_pred= rfr.predict(X_test)
```

In [47]:
```
rfr.feature_importances_
```

Out[47]:
```
array([0.07211201, 0.33454801, 0.08115789, 0.0491683 , 0.05340307,
       0.17098935, 0.12161647, 0.1170049 ])
```

In [48]:
```
mae = mean_absolute_error(y_test, y_pred)
mse = mean_squared_error(y_test, y_pred)
r2 = r2_score(y_test, y_pred)

mae, mse, r2
```

Out[48]: (0.3118954248366013, 0.16183790849673205, 0.312438548094374)