

Regressions on Fertility Data Implemented Using Scikit-Learn

Preprocessing data

```
In [ ]: with open('fertility.txt') as fertility:
    patients = []
    classification = []
    for line in fertility:
        data = line.split()
        for i in data:
            ndata = i.split(',')
            patients += [ndata]
    for p in range(len(patients)):
        classification += patients[p][-1]
        patients[p] = patients[p][: -1]
    for j in patients:
        for k in range(len(j)):
            j[k] = float(j[k])
    for l in range(len(classification)):
        if classification[l] == 'N':
            classification[l] = 0
        else:
            classification[l] = 1
    final = [patients, classification]
```

Linear Regression

```
In [ ]: from sklearn import linear_model
reg = linear_model.LinearRegression()
reg.fit(final[0], final[1])
print(reg.coef_)
```

Decision Tree

```
In [ ]: from sklearn import tree
import graphviz
import matplotlib.pyplot as plt
x = final[0]
y = final[1]
clf = tree.DecisionTreeRegressor()
clf = clf.fit(x, y)
graph = tree.export_text(clf)
print(graph)
```