

PREDICTIVE MODELING FOR DIABETIC RETINA USING MACHINE LEARNING ALGORITHMS

Team Members:
Mohammed Salmaan.S
Md Faisal.A

Note

Datasets is downloaded from UCI-Repository

import the destination of the retina file here, for me its in C://Users/SALMAAN/Documents/retina.csv

Rows and Columns

```
In [1]: from pandas import read_csv
#import the destination of the retina file here, for me its in C://Users/SALMAAN/Documents/retina.csv
filename = 'C://Users/SALMAAN/Documents/retina.csv'
dataframe = read_csv(filename)
rc=dataframe.shape
print(rc)

(1151, 20)
```

First Two Rows

```
In [2]: from pandas import read_csv
#import the destination of the retina file here, for me its in C://Users/SALMAAN/Documents/retina.csv
filename = 'C://Users/SALMAAN/Documents/retina.csv'
two=dataframe.head(2)
print(two)
```

	0	1	2	3	4	5	6	7	8	9	10	11	\
0	1	1	22	22	22	19	18	14	49.895756	17.775994	5.270920	0.771761	
1	1	1	24	24	22	18	16	13	57.709936	23.799994	3.325423	0.234185	
	12	13	14	15	16	17	18	Class					
0	0.018632	0.006864	0.003923	0.003923	0.486903	0.100025	1	0					
1	0.003903	0.003903	0.003903	0.003903	0.520908	0.144414	0	0					

Confusion Matrix

```
In [3]: from pandas import read_csv
from sklearn.metrics import confusion_matrix
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
#import the destination of the retina file here, for me its in C://Users/SALMAAN/Documents/retina.csv
filename = 'C://Users/SALMAAN/Documents/retina.csv'
array = dataframe.values
X = array[:,0:19]
Y = array[:,19]
test_size = 0.33
seed = 7
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=test_size, random_state=seed)
model = LogisticRegression()
model.fit(X_train, Y_train)
predicted = model.predict(X_test)
matrix = confusion_matrix(Y_test, predicted)
print(matrix)

[[145  25]
 [ 70 140]]
```

Logistic Regression

```
In [4]: from pandas import read_csv
from sklearn.metrics import confusion_matrix
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.model_selection import KFold
from sklearn.model_selection import cross_val_score
#import the destination of the retina file here, for me its in C://Users/SALMAAN/Documents/retina.csv
filename = 'C://Users/SALMAAN/Documents/retina.csv'
array = dataframe.values
X = array[:,0:19]
Y = array[:,19]
validation_size = 0.20
seed = 7
models=[]
models.append(('LR', LogisticRegression()))
results = []
names = []
for name, model in models:
    kfold = KFold(n_splits=10, random_state=seed)
    cv_results = cross_val_score(model, X, Y, cv=kfold, scoring='accuracy')
    results.append(cv_results)
    names.append(name)
    msg = "%s: %f (%f)" % (name, cv_results.mean(), cv_results.std())
    print(msg)

LR: 0.748951 (0.032067)
```

Linear Discriminant Analysis

```
In [5]: from pandas import read_csv
from sklearn.metrics import confusion_matrix
from sklearn.model_selection import train_test_split
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.model_selection import KFold
from sklearn.model_selection import cross_val_score
#import the destination of the retina file here, for me its in C://Users/SALMAAN/Documents/retina.csv
filename = 'C://Users/SALMAAN/Documents/retina.csv'
array = dataframe.values
X = array[:,0:19]
Y = array[:,19]
validation_size = 0.20
seed = 7
models=[]
models.append(('LR', LinearDiscriminantAnalysis()))
results = []
names = []
for name, model in models:
    kfold = KFold(n_splits=10, random_state=seed)
    cv_results = cross_val_score(model, X, Y, cv=kfold, scoring='accuracy')
    results.append(cv_results)
    names.append(name)
    msg = "%s: %f (%f)" % (name, cv_results.mean(), cv_results.std())
    print(msg)
```

LR: 0.722001 (0.028613)

Decision Tree

```
In [6]: from pandas import read_csv
from sklearn.metrics import confusion_matrix
from sklearn.model_selection import train_test_split
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import KFold
from sklearn.model_selection import cross_val_score
#import the destination of the retina file here, for me its in C://Users/SALMAAN/Documents/retina.csv
filename = 'C://Users/SALMAAN/Documents/retina.csv'
array = dataframe.values
X = array[:,0:19]
Y = array[:,19]
validation_size = 0.20
seed = 7
models=[]
models.append(('LR', DecisionTreeClassifier()))
results = []
names = []
for name, model in models:
    kfold = KFold(n_splits=10, random_state=seed)
    cv_results = cross_val_score(model, X, Y, cv=kfold, scoring='accuracy')
    results.append(cv_results)
    names.append(name)
    msg = "%s: %f (%f)" % (name, cv_results.mean(), cv_results.std())
    print(msg)
```

LR: 0.608171 (0.037535)

Algorithm Comparison

```

In [7]: from pandas import read_csv
        from sklearn.metrics import confusion_matrix
        from sklearn.model_selection import train_test_split
        from sklearn.linear_model import LogisticRegression
        from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
        from sklearn.tree import DecisionTreeClassifier
        from sklearn.model_selection import KFold
        from sklearn.model_selection import cross_val_score
        #import the destination of the retina file here, for me its in C://Users/SALMAAN/Documents/retina.csv
        filename = 'C://Users/SALMAAN/Documents/retina.csv'
        array = dataframe.values
        X = array[:,0:19]
        Y = array[:,19]
        validation_size = 0.20
        seed = 7
        models = []
        models.append(('LR', LogisticRegression()))
        models.append(('LDA', LinearDiscriminantAnalysis()))
        models.append(('CART', DecisionTreeClassifier()))
        results = []
        names = []
        for name, model in models:
            kfold = KFold(n_splits=10, random_state=seed)
            cv_results = cross_val_score(model, X, Y, cv=kfold, scoring='accuracy')
            results.append(cv_results)
            names.append(name)
            msg = "%s: %f (%f)" % (name, cv_results.mean(), cv_results.std())
            print("-----ALGORITHM---COMPARISON-----")
            print(msg)

-----ALGORITHM---COMPARISON-----
LR: 0.748951 (0.032067)
-----ALGORITHM---COMPARISON-----
LDA: 0.722001 (0.028613)
-----ALGORITHM---COMPARISON-----
CART: 0.603831 (0.026860)

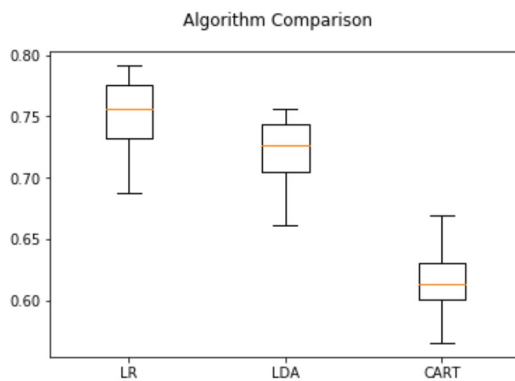
```

Comparison Plotting

```
In [8]: from pandas import read_csv
from sklearn.metrics import confusion_matrix
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.tree import DecisionTreeClassifier
from sklearn.model_selection import KFold
from sklearn.model_selection import cross_val_score
from matplotlib import pyplot
#import the destination of the retina file here, for me its in C://Users/SALMAAN/Documents/retina.csv
filename = 'C://Users/SALMAAN/Documents/retina.csv'
array = dataframe.values
X = array[:,0:19]
Y = array[:,19]
validation_size = 0.20
seed = 7
models = []
models.append(('LR', LogisticRegression()))
models.append(('LDA', LinearDiscriminantAnalysis()))
models.append(('CART', DecisionTreeClassifier()))
results = []
names = []
for name, model in models:
    kfold = KFold(n_splits=10, random_state=seed)
    cv_results = cross_val_score(model, X, Y, cv=kfold, scoring='accuracy')
    results.append(cv_results)
    names.append(name)
    msg = "%s: %f (%f)" % (name, cv_results.mean(), cv_results.std())
    fig = pyplot.figure()
fig.suptitle('Algorithm Comparison')
ax = fig.add_subplot(111)
pyplot.boxplot(results)
ax.set_xticklabels(names)
pyplot.show()
```

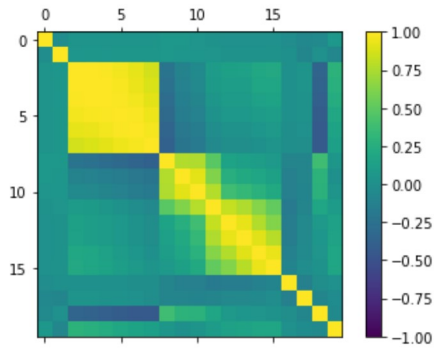
<matplotlib.figure.Figure at 0x266c0266780>

<matplotlib.figure.Figure at 0x266c05db9b0>



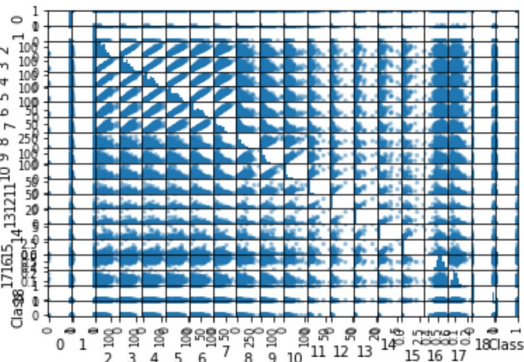
Correlation Matrix

```
In [9]: from pandas import read_csv
from pandas.plotting import scatter_matrix
from matplotlib import pyplot
#import the destination of the retina file here, for me its in C://Users/SALMAAN/Documents/retina.csv
filename = 'C://Users/SALMAAN/Documents/retina.csv'
correlations = dataframe.corr()
fig = pyplot.figure()
ax = fig.add_subplot(111)
cax = ax.matshow(correlations, vmin=-1, vmax=1)
fig.colorbar(cax)
pyplot.show()
```



Scatter Matrix

```
In [10]: from pandas import read_csv
from pandas.plotting import scatter_matrix
from matplotlib import pyplot
#import the destination of the retina file here, for me its in C://Users/SALMAAN/Documents/retina.csv
filename = 'C://Users/SALMAAN/Documents/retina.csv'
scatter_matrix(dataframe)
pyplot.show()
```



Making Predictions With Input

Diabetic Retinopathy Postive Detection

```
In [11]: import pandas as pd
import numpy as np
from pandas import read_csv
from sklearn.model_selection import train_test_split
from sklearn.model_selection import KFold
from sklearn.model_selection import cross_val_score
from sklearn.linear_model import LogisticRegression
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report
from pandas.plotting import scatter_matrix
from sklearn.metrics import accuracy_score
from matplotlib import pyplot

#import the destination of the retina file here, for me its in C://Users/SALMAAN/Documents/retina.csv
filename = 'C://Users/SALMAAN/Documents/retina.csv'
dataframe = read_csv(filename)

array = dataframe.values
X = array[:,0:19]
Y = array[:,19]
test_size = 0.25
seed = 7
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=test_size,
random_state=seed)
model = LogisticRegression()
model.fit(X_train, Y_train)
predicted = model.predict(X_test)
matrix = confusion_matrix(Y_test,predicted)
testing=model.predict([[1,1, 62, 60,59 ,54, 47,33, 55,27, 12, 4,1,0, 0, 0, 0, 0, 0 ]])
print (testing)

[ 1.]
```

Diabetic Retinopathy Negative Detection

```
In [12]: import pandas as pd
import numpy as np
from pandas import read_csv
from sklearn.model_selection import train_test_split
from sklearn.model_selection import KFold
from sklearn.model_selection import cross_val_score
from sklearn.linear_model import LogisticRegression
from sklearn.discriminant_analysis import LinearDiscriminantAnalysis
from sklearn.tree import DecisionTreeClassifier
from sklearn.metrics import confusion_matrix
from sklearn.metrics import classification_report
from pandas.plotting import scatter_matrix
from sklearn.metrics import accuracy_score
from matplotlib import pyplot

#import the destination of the retina file here, for me its in C://Users/SALMAAN/Documents/retina.csv
filename = 'C://Users/SALMAAN/Documents/retina.csv'
dataframe = read_csv(filename)

array = dataframe.values
X = array[:,0:19]
Y = array[:,19]
test_size = 0.25
seed = 7
X_train, X_test, Y_train, Y_test = train_test_split(X, Y, test_size=test_size,
random_state=seed)
model = LogisticRegression()
model.fit(X_train, Y_train)
predicted = model.predict(X_test)
matrix = confusion_matrix(Y_test,predicted)
testing=model.predict([[1,1, 22, 22,22 ,19, 18, 14, 49, 17, 5, 0, 0, 0, 0, 0, 0, 0, 1 ]])
print (testing)

[ 0.]
```