import pandas as pd

filename = 'cassava dataset.xlsx'

df = pd.read\_excel(filename)

df.rename(columns={'total leaf length (cm)': 'leaf\_length', 'stem diameter (cm)': 'stem\_diameter', 'Yield (kg)': 'yield'}, inplace=True)

df.shape

corr = df.corr()

corr.style.background\_gradient(cmap='BrBG')

import seaborn as sns

sns.heatmap(df.corr())

sns.pairplot(df)

import matplotlib.pyplot as plt

plt.hist(df['yield'], bins=40)

from sklearn.model\_selection import train\_test\_split

X = df.drop('yield', axis=1)

y = df['yield']

X\_train, X\_val, y\_train, y\_val = train\_test\_split(X, y, random\_state=11, test\_size=0.2)

X\_val.shape

from sklearn.ensemble import RandomForestRegressor

from sklearn.neighbors import KNeighborsRegressor

from sklearn.linear\_model import LinearRegression

model = RandomForestRegressor()

model.fit(X\_train, y\_train)

model.score(X\_val, y\_val)

model.predict(X\_val)

y\_val.values

dummy = X\_val.copy()

dummy['correct\_label'] = y\_val

dummy['predicted\_label'] = model.predict(X\_val)

dummy

LINK TO THE GOOGLE COLAB WORK SPACE WHERE THE WORK WAS DONE

[https://colab.research.google.com/drive/1CSaRDpOHLSfqa3FaSXYbFuhWQMcc-C1B#scrollTo=neck\_XSkE9ln](codes%20.docx)