Anthony Camacci

INFO 3401-001

Problem set 3

Worked with Spencer, Ji, Steven, Skyler, Ryan

1. I think this data is for seeing the likelihood of someone returning a loan and what factors can

impact this.

1. def loadAndCleanData(fName):

data = pd.read\_csv(fNAME)

nData = data.fillna(“0”)

return(nData)

1. from utils import \*

df = loadAndcleanData(“creditData.csv”)

1. def computePDF(columnName, DS):

nPlot = DS[columnName].plot.kde()

plt.show(nPLot)

1. def viewDistribution(CName, DS):

nPlot = dataset.hist(column = cName)

plt.show(nPlot)

1. def viewLogDistribution(cName, DS):

nPlot = DS.hist(column = cName, log = True)

plt.show(nPlot)

1. def binsCount(cNAme, data):

bins = pd.qcut(data[cName], q = 3, duplicate = ‘drop’)

print(bins[0])

cNames = list(cleanedData.columns.values)

for I in columnNames

newBins = bins(I, cleanedData)

1. def computeDefaultRisk(column, Tf, bin, DF):

total = ‘’

total2 = ‘’

for I, datapoint in dataframe.iterrows():

if datapoint[Tf]>=bin[0] and datapoint[Tf]<bin[1]:

count+=1

if dataPoints[column]==1

count2 +=1

lengthData = len(data)

prob = total/lengthData

prob2 = total/lengthData

return(prob2/prob)

1. ComputeDefaultRisk
2. loadAndCleanData(“newLoans.csv”)
3. ‘Age’: 0.025, ‘Number of Dependents’: 0.025, ‘Monthly Income’: 0.010, ‘Debt Ratio’: 0.010, ‘Revolving Utilization’: 0.010, ‘Number of Credit Lines & Loans ’: 0.010, ‘Number of Real Estate Loans or Lines’: 0.010, ‘Number of Times 30-59 Days Late’: 0.015, ‘Number of Times 60-89 Days Late’: 0.015, ‘Number of Times 90 Days Late’: 0.015
4. For row in range(len(newLoans.index)):

Val = predictDefaultRisk(newLoans.iloc[[row]],risks,weights)

newLoans[‘SeriousDlqin2yrs’][row] = val

1. CopmputePDF(‘Seriousdlqin2yrs’,df)