

Texas Tech University
Department of Computer Science

Course Name: Introduction to Artificial Intelligence **Number:** CS3368 **Period:** Fall 2025
Instructor: Dr. Juan Carlos Rojas **Email:** Juan-Carlos.Rojas@ttu.edu

Homework 4

Submit your programs packed together into a single ZIP file, and a single report document in Word or PDF that shows your programs working. For the report include screen shots, or copy-paste of the terminal shell that shows the commands used and the output produced.

Problem 1

This problem focuses on the Credit Delinquency dataset. It attempts to make better use of the credit record histories to extract additional input features, with the hope of improving the quality of the delinquency predictor.

Part A

Create a Python program similar to `Credit_CreateLabels.py`, with the following modifications:

- All entries with a Status of X should be ignored. This is justified because customers with no loans are neither “good” or “bad” at paying their loans and should not be mixed with customers with loans.
 - Note that this may eliminate some customers from the dataset, those who only had statuses of X.
- Compute the total counts of different Status values for each customer. We are interested in the total number of “C”, “0” and “1” status, as well as the total observations (not including Xs).
- Calculate the fraction of the observations that are “C”, “0” and “1” for each customer. Store them as new columns in the table.
- Compute and store the “Delinquent” column as before: “1” if the customer has any number of statuses equal to “2”, “3”, “4” or “5”, and zero otherwise.

Store the new data into a new file called “credit_delinquency_v2.csv”.

Part B

Explore the characteristics of the new columns.

- Show histograms of the “C”, “0” and “1” ratio columns.
- Show bar plots of the delinquency rate for customers in different intervals of these ratio columns.

Part C

Prepare the data for prediction as usual:

- Expand the categorical columns into dummy one-hot encoded columns
- Separate the target column (“Delinquent”) from the remaining data columns
- Shuffle and split into training and test subsets, using the random state 2025.
- Standardize the scale of all input columns

Create a predictor for delinquency using Logistic Regression. Do not do employ any class balancing strategy.

- Compute the Precision/Recall graph
- Compute the Receiver Operating Characteristic graph
- Compute the ROC AUC metric

Contrast the result of this predictor with the example seen in class (with `credit_delinquency.csv`).

- Did the new input columns help improve the predictions?

Part D

Determine the importance of the new input columns in the predictor of part C, by comparing their coefficient magnitudes with others in the model.

- How important are each of the new columns to your predictor?
- Does this correlate with the improvement of the results?