Alessandra Bielli and Aiden Hennessy
Stat 627 - Statistical Machine Learning
Dr. Richard Ressler
10 December 2024

For the final project, we assessed how big the racial and ethnic home lending approval disparities are in Mecklenburg County, North Carolina. We decided on this topic together and found the Mecklenburg County Loan Application dataset through the Consumer Financial Protection Bureau's website. Once, we obtained the data, we came up with two questions: for regression, "How do an applicant's income, race, and credit score (the race of the primary applicant) predict their loan-to-value (LTV) ratio?" and for classification, "Can we predict whether a loan application will be approved or denied (action taken) based on factors like income, race, debt-to-income ratio, credit score, and loan-to-value ratio?".

Next, we created a project plan that outlined our goals: composed plan (week 10), cleaning and preparation (week 11), data visualization (week 12), data analysis (week 13), data presentation prep, and report (week 14).

As a result, each of us decided to tackle one of these questions. Alessandra took the regression question and analyzed the data using Linear Regression, Ridge Regression, and Partial Least Squares Regression. This involved cleaning the data for analysis, splitting the data into 70% training and 30% testing, fitting the models, and comparing the results.

Aidan took the classification question and utilized LDA, KNN, and Tree. He also created some Anova functions, which showed that the response variable (action taken) is easier to predict when predictor variables about race and ethnicity are included than when predictor variables about race and ethnicity are not included.

We met a total of 19 times beginning on October 18th. These meetings included progress checks, code evaluations, and project deliverable collaborations (e.g. the final poster).

Leading up to the final presentation, we shared our findings and worked together to create a poster that reflected this. We also made a GitHub repository to store their completed code files, the data used for their analysis, and a readme file containing the necessary information to reproduce their findings.