STATS 771 - Project Proposal
Applying KCV-SMOTE to Linear Regression

Pao Zhu Vivian Hsu (Student Number: 400547994)

Invalid Date

Introduction

- Description about credit card fraud and how it impacts companies and customers
- Mention that this paper will do KCV-SMOTE using a different classifier and compare results with the paper (which uses the same dataset as)

Literature Review

- Description of the the proposed method (and how they suggested a method to handle)
- Description of any advancements since the paper
-

Data and Methods

The proposed study will use public data on credit card fraud from the Kaggle website [].

Expected Results

We expect the results to perform better than linear regression model and the normal classifier.

Timelines

This study will contain two components, a written report and a presentation. In this section, we describe targeted dates for completing both components. We aim to finish data collection, preliminary visualization, and data splitting into training and testing sets by November 30, 2023. By December 31, 2023, we will finish the modelling, including a model for the default classifier method and another model using KCV-SMOTE version. By January 31, 2024, we will assess each model's performance and compare the results between models. The results will also be compared with Kang and Zhang's study (2022). By February 28, 2024, we will finish a draft of the report. By March 31, 2024, we will prepare the presentation slides. By April 20, 2024, the report and presentation will be finalized. The presentation will occur during the last week of April.

References

Kang, H., & Zhang, H. (2022). A new improved method for online credit anti-fraud. *Automatic Control and Computer Sciences*, *56*, 347–355. https://doi.org/10.3103/S0146411622040046

R Core Team. (2023). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/