

Introduction

The utilization of statistics, if performed effectively, can be the deciding factor as to whether a credit lender is successful. While the approval of credit for a consumer who will pay on time results in \$250 in revenue on average, the approval of credit for a consumer who later defaults can be a particularly costly mistake leading to approximately \$750 in losses on average per instance. As a result, the successful identification of consumers who will not default is a priority for credit lenders.

With this in mind, we sought to create a model that could be utilized to effectively predict whether consumers are likely to default. The data was gathered by from a major credit bureau and contained 1,462,955 observations of consumers prior to approval. Another data set contained the 17,244,104 observations for the longitudinal post-hoc performance of each consumer after approval.