

Credit Card Fraud

Credit card fraud costs consumers an estimated \$22 billion per year. Due to the importance of this issue, credit card companies typically have teams of data scientists specifically tasked with identifying fraudulent transactions. The dataset CCFraud.csv contains information on about 300,000 credit card transactions of which only 492 are known to be fraudulent (about 0.1%). The dataset also contains information on 29 different characteristics of the transaction. However, because such information is highly proprietary and confidential, most of the explanatory variables (labeled simply as V1-V28) are the principal component scores of the explanatory variables. Your task is to try to identify which transactions are fraudulent and present it to the legal team at your company. In your report, please address the following questions:

1. How accurately can you identify the fraudulent transactions? In other words, given a transaction is fraudulent, how well do you identify it as such?
2. The file IsFraudulent.csv contains information on 5 different transactions. Which of these transactions, if any, do you think are fraudulent?