TECNOLÓGICO DE MONTERREY

DATA ANALYTICS BOOTCAMP

MODULE 17 – SUPERVISED MACHINE LEARNING AND CREDIT RISK

---

1. OVERVIEW OF THE ANALYSIS

Fast Loan, a peer-to-peer lending firm, intends to anticipate credit risk using machine learning. Management believes that this will result in a more efficient and dependable loan process. It also believes that Machine Learning will improve the accuracy of identifying qualified loan candidates, resulting in decreased default rates.

The firm wants me to help the lead Data Scientist put this idea into action. Fast Lending, a peer-to-peer lending services startup, wants to employ machine learning to forecast credit risk, and I'll be building and evaluating many models in my job. Management believes that this will result in a more efficient and dependable loan process. It also believes that machine learning will aid in the more accurate identification of strong credit risk models or algorithms. To get the most out of your models and data, I'll utilize techniques like resampling and boosting. I'll test their performance and see how well your models predict data once I've built and implemented these techniques.

I'll dig headfirst into Machine Learning algorithms, statistics, and data processing approaches to complete my objective.

PURPOSE: Build and evaluate several Machine Learning models or algorithms to predict credit risk, using techniques such as re sampling and boosting, and evaluate their performance to see how well your models predict data.

2. RESULTS OF THE ANALYSIS

Table 1. Consolidated results that describe the balanced accuracy score and the precision and recall scores of all six machine learning models

MACHINE LEARNING MODEL	BALANCED ACCURACY SCORE	PRECISION	RECALL
Naive Random Oversampling	0.8443807716	0.84	0.84
SMOTE Oversampling	0.8828795124	0.88	0.88
Undersampling	0.9208494208	0.92	0.92
Combination (Over and Under) Sampling	0.8768478321	0.88	0.88
Balanced Random Forest Classifier	0.7587927757	0.99	0.89
Easy Ensemble AdaBoost Classifier	0.917333617	0.99	0.95

3. SUMMARY OF THE ANALYSIS

The first five models all provide comparable results and have an accuracy of roughly 75% to 92%. The Easy Ensambles Classifier is the apparent standout, with a 93.17 percent accuracy and 92 percent sensitivity, indicating that we are catching the majority of high-risk loans. This model is the apparent recommendation, however it may be necessary to double-check if the high-risk loans are indeed high-risk, or if they are simply model faults.