
HERRAMIENTAS DE MINERÍA DE DATOS PARA EL MODELAMIENTO DE LA PÉRDIDA DADO EL DEFAULT

JUAN FRANCISCO TOLEDO LOYOLA
MAGISTER EN GESTIÓN DE OPERACIONES

RESUMEN

Esta tesis aborda el problema de la Pérdida dado el Default (LGD, por sus siglas en inglés), utilizando diferentes herramientas de minería de datos para su medición.

El problema nace de las dificultades que se encuentran en la estimación de las provisiones para las entidades financieras, ya que dependen de parámetros como la EAD o Exposición al Default que corresponde al monto total que el cliente le debe a la compañía financiera, por lo cual es un valor conocido, otro parámetro importante es la PD o Probabilidad de Default que corresponde a la posibilidad de que un cliente devuelva lo solicitado, medición realizado principalmente por modelos de scoring, y por último se tiene a la LGD o Pérdida dado el Default que corresponde a la pérdida esperada para la entidad financiera en caso de que ocurra el default, la cual posee poca incursión en el modelamiento.

Es necesario utilizar minería de datos para abordar este problema, ya que se pueden establecer predicciones en base al análisis de grandes bases de datos, considerando variables sobre características crediticias históricas del cliente y de indicadores macroeconómicos del país, al momento de que un cliente caiga en default. Para lo anterior se utilizó la metodología de descubrimiento de conocimiento en base de datos (KDD) para la construcción tanto de los modelos propuestos como de los modelos a comparar.

Los métodos implementados corresponden a modelos no lineales, como lo son Support Vector Regression (SVR), Ensembles y SVR con programación de Kernels, estos modelos son comparados con la literatura actual, en donde destacan modelos de Redes Neuronales, Árboles de Regresión y Regresión con respuesta fraccional.

Se realiza una comparación mediante una aplicación a un caso real, en donde las propuestas entorno a la metodología ensemble son consideradas como las más competentes, ya que a lo menos igualan los mejores resultados actuales

obtenidos por las Redes Neuronales, siendo el Random Forest una atractiva idea de modelar la Pérdida dado el Default.

Palabras claves: Riesgo Crediticio, Pérdida dado el Default, SVR, ensembles.

ABSTRACT

This thesis addresses the problem of Loss given Default (LGD, for its acronym in English), using different data mining tools for measurement.

The problem stems from the difficulties encountered in estimating provisions for financial institutions, as they depend on parameters like EAD or Exposure to the Default corresponding to the total amount that the customer owes the finance company, so is a known value, another important parameter is the PD or default probability corresponding to the likelihood that a customer returns the requested measurement performed mainly by scoring models, and finally has LGD or Loss Given Default that corresponds to the expected loss for the lender in the event of default, which has little foray into modeling.

You need to use data mining to address this problem, since they can make predictions based on the analysis of large databases, considering variables on historical customer credit characteristics and macroeconomic indicators of the country, when a customer falls in default. For the above, the methodology of knowledge discovery in databases (KDD) for construction of the models proposed so as to compare models.

The methods are implemented to nonlinear models, such as Support Vector Regression (SVR), and SVR Ensembles Kernels with programming, these models are compared with the current literature, where models stand Neural Networks, Regression and Regression Trees with fractional response.

A comparison is made by an application to a real case, where the methodology proposed ensemble environment are considered the most relevant, since at least match the best current results obtained by neural networks, with the Random Forest an attractive idea of modeling the Loss given

Default. Keywords: Credit Risk, Loss given Default, SVR ensembles.