Bankruptcy prediction in Colombian case, using multilayer perceptron trained with memetic algorithm

Student:Iván Andrés Trujillo Abella,
Proffesors: Eliana María González Neira & Gabriel Mauricio Zambrano Rey

***Abstract*—Literature about Bankruptcy prediction is still incipient, therefore this work try to fill this gap by using machine learning and metaheuristics techniques to find an optimal set of weights in a MLP model.**

***Index Terms*—Machine learning, Bankruptcy, Metaheuristics, Evolutionary Algorithms, Local Search, Memetic algorithms, Neural Networks, Multilayer Perceptron.**

I. Information(Data)

Data was retrieved from SIS by the period 2016-2019, classifying as bankrupt all Small and Medium Enterprises(SMEs) firms that enter in process of insolvency.

A. Event definition

The event was determined as firms that enter of one year to another in a insolvency process.

To model the phenomena were used financial ratios of one and two year before of the occurrence of the event. For instance, those firms that have a normal state in 2018 and enter in insolvency process in 2019 the financial ratios of 2018 and 2017 to model the event in 2019.

For those firms that not present the event(controls) were used the mean of financial ratio by its period of activity. For instance, j firm enter to the database in 2017 and not present the event then were used the mean of the financial ratios for the period (2017 - 2019).

B. Exclusion criteria

- Were excluded financial information present in a date different to December 31 of each year.
- Were excluded from database firms that declared in a preoperative condition each year (this and the former criteria give us unique firms).
- Were excluded from database those firms that present the event for the initial year (2016).
- Were excluded firms that present missing values in financial ratios using to prediction.

II. Description of data

To describe categorical data were used absolute and relative frequencies. The numerical variables were describe with mean and standard deviation if the variable is normally distributed, otherwise were used median and 25% and 75% percentile.

To compare the financial ratios and determine if there are significance difference among bankrupt and no-bankrupt firms, we used t-test or wilcoxon rank sum test if the variable is normal or not respectively, to test normally we used shapiro-wilk test. To determine if there are independence among the event and categorical variables the χ^2 test was used.

III. Modelling prediction

The models used to benchmark were: Decision tree, logistic regression and multilayer perceptron.

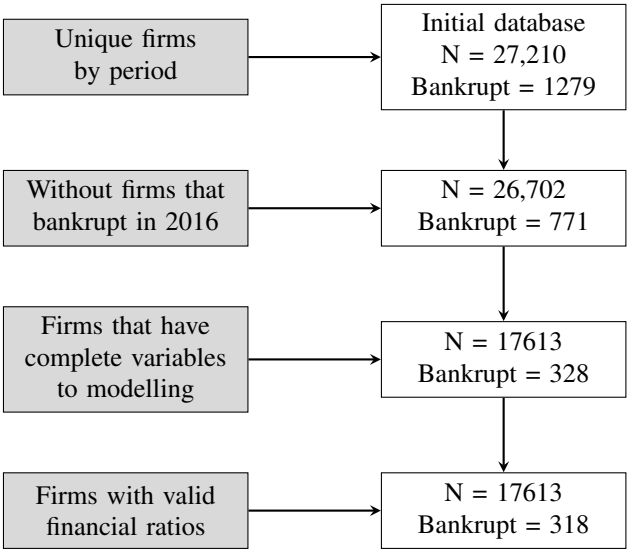
A. Train and test

We used 65% to train and 35% to test.

B. Balance data

We used udersampling.

IV. Results



For the period (2016-2019) 700 firms were found as

According to the following figure the spearman coefficient show the monotocity relationship among the predictors.

There are a relationship among the variables.

V. Appendix

The financial ratios used were:

- new

were

TABLE I: Model performance with one year of lag

	Logistic Regression		Decision Tree		Multilayer Perceptron	
	Default	No-deafult	Default	No-default	No-default	Default
precision	1.00	0.53	0.79	0.58	0.00	0.49
recall	0.12	1.00	0.38	0.90	0.00	1.00
f1-score	0.22	0.69	0.51	0.71	0.00	0.66
support	40.00	39.00	40.00	39.00	40.00	39.00

TABLE I: Model performance with one year of lag

n	
time-event, n (%)	0.0 2017.0 2018.0 2019.0
MGB, median [Q1,Q3]	
MGN, median [Q1,Q3]	
ROE, median [Q1,Q3]	
ROA, median [Q1,Q3]	
NE, median [Q1,Q3]	
PCP, median [Q1,Q3]	
Ax1, median [Q1,Q3]	
Ax2, median [Q1,Q3]	
Sector, n (%)	Actividades artísticas, de entretenimiento y recreación Actividades de atención de la salud humana y de asistencia social Actividades de organizaciones y entidades extraterritoriales Actividades de servicios administrativos y de apoyo Actividades financieras y de seguros Actividades inmobiliarias Actividades profesionales, científicas y técnicas Administración pública y defensa; planes de seguridad social de afiliación obligatoria Agricultura, ganadería, caza, silvicultura y pesca Alojamiento y servicios de comida Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas Construcción Distribución de agua; evacuación y tratamiento de aguas residuales, gestión de desechos y actividades de saneanmiento ambiental Educación Explotación de minas y canteras Industrias manufacturetas Información y comunicaciones Otras actividades de servicios Suministro de electricidad, gas, vapor y aire acondicionado Transporte y almacenamiento

Fig. 1: Spearman correlation in predictor variables