Interdisciplinary Congress of Renewable Energies - Industrial Maintenance - Mechatronics and Informatics
Booklets

RENIECYT - LATINDEX - Research Gate - DULCINEA - CLASE - Sudoc - HISPANA - SHERPA UNIVERSIA - Google Scholar DOI - REDIB - Mendeley - DIALNET - ROAD - ORCID

Title: Neural networks to predict academic dropout in engineering

Authors: CENTURIÓN-CARDEÑA, Humberto José, CANO-BARRÓN, Danice Deyanira, SANDOVAL-GÍO, Jesús y ZAPATA-GONZÁLEZ, Alfredo

Editorial label ECORFAN: 607-8695
BCIERMMI Control Number: 2020-04
BCIERMMI Classification (2020): 211020-0004

Pages: 14
RNA: 03-2010-032610115700-14

ECORFAN-México, S.C.
143 – 50 Itzopan Street
La Florida, Ecatepec Municipality
Mexico State, 55120 Zipcode
Phone: +52 1 55 6159 2296
Skype: ecorfan-mexico.s.c.
E-mail: contacto@ecorfan.org
Facebook: ECORFAN-México S. C.
Twitter: @EcorfanC

www.ecorfan.org

Holdings
Mexico Colombia Guatemala
Bolivia Cameroon Democratic
Spain El Salvador Republic
Ecuador Taiwan of Congo
Peru Paraguay Nicaragua

Introduction

Methodology

Results

Conclusions

References

Introduction

The strategies oriented to efficiently storage of data and search not evident patterns have become an important area in computer science (Riquelme, Ruiz, & Gilbert, 2006)

Educational Data Mining is a paradigm to design models, tasks, methods and algorithms to explore data from educational environments to find patterns and make predictions of the behavior and achievement of students, evaluation of functionality and educational applications in conventional, open and remote environments (Luan, 2002)

Introduction

Predicting student success is crucial for higher education because the quality of the teaching and learning process is strongly related to the their ability to respond to students needs (Al-Twijri & Noamanb, 2015)

This study focuses on the analysis of context data and the results of the EXANI-II skills assessment, which is a standard instrument used in Mexico to identify the academic and personal status of higher education applicants from 2015 to 2018 from Motul Higher Technological Institute

Methodology

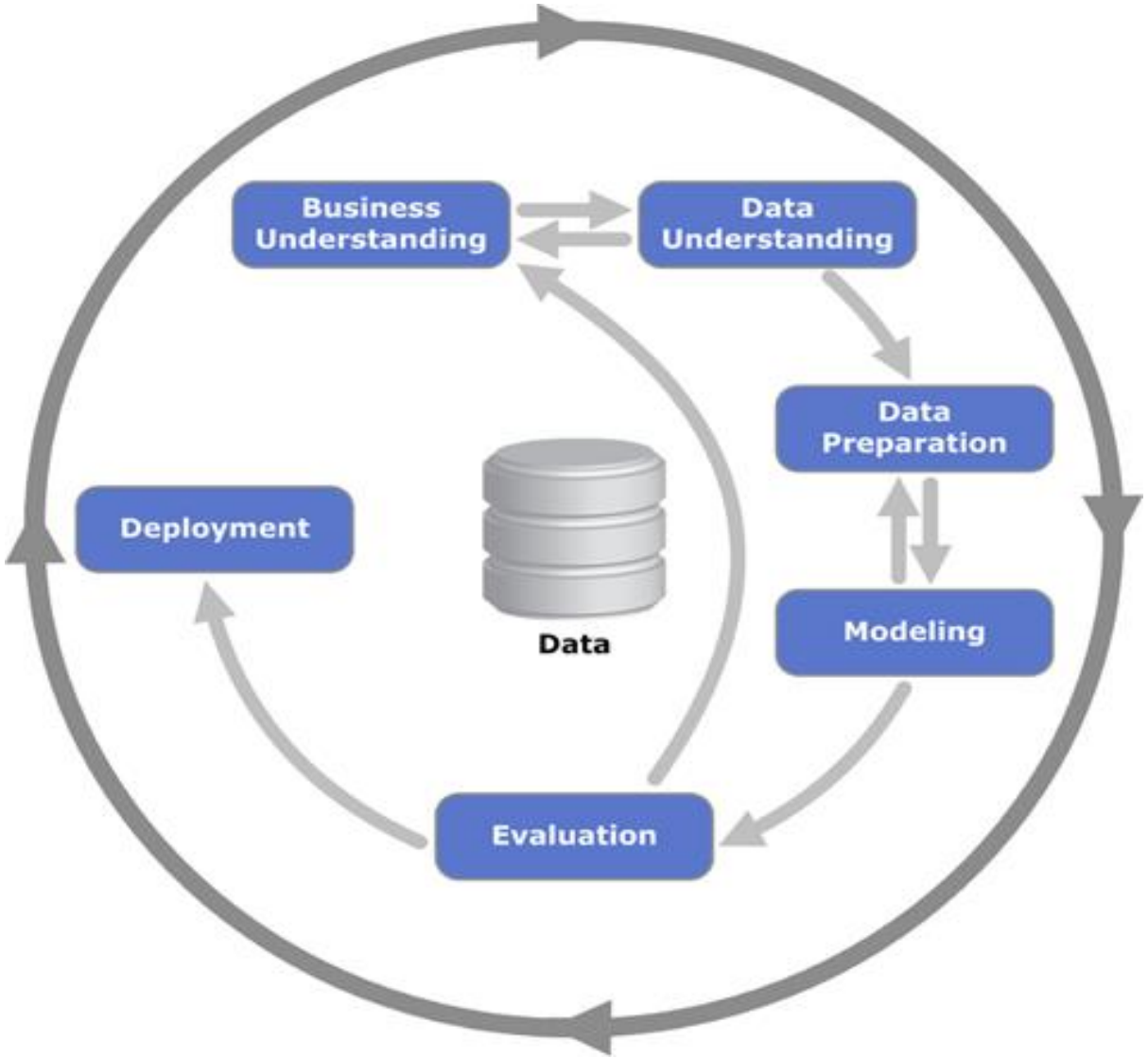

Figura 1 CRISP-DM life cicle.
Fuente: (Chapman, Khabza, & Shearer, 2000)

Methodology

ID sources of information in the organization and the type of files that are available to use

Join all the different useful data bases identified (Python)

To avoid overfitting and underfitting select the variables with more significance to the independent variable

Select the neural network model according to thumb rules

Evaluate the prediction efficiency of the models and select the best model for deployment

Methodology

Neural network design process:

- The number of layers considered are one or two, one can approximate any function that maps continuously from one finite space to another, while two can approximate any mapping with a certain level of efficiency (Heaton,2008)
- Regarding the number of neurons, the rule of the average number of inputs plus the output was considered
- The sigmoid or logistic function was used since its outputs values range between 0 and 1

Methodology

The principal component analysis (PCA) aims to replace redundant attributes with new attributes that adequately integrate the information originally contained

Atributo	Clave	Importancia
1	rezago	81.4
2	prom_bac	14.2
3	dan_reqc	8.9
4	fecha_apli	7.5
5	ipan	7.5
6	ipma	7.2
7	iele	6.9
8	dan_malf	6.3
9	dan_eir	6
10	dan_ofi	5.9
11	icne	5.7
12	fre_cde	3.8
13	vac_rm	3.5
14	fre_tsc	3.4
15	dan_mft	3.3
16	icle	3.3
17	edad_ingreso	3
18	fre_sme	2.5
19	hrs_trab	2.5
20	ser	2.1

Methodology

Tasa de aprendizaje	Atributos de entrada	Mínimo de neuronas en la capa oculta
.01	20	11
	15	8
	10	6
.001	20	11
	15	8
	10	6
.05	20	11
	15	8
	10	6

Results

Once the parameters for the tests were decided, the databases were generated and the learning process was run, with a test database of 15% of the original database whose records were selected randomly

Results

Tasa de aprendizaje	Atributos de entrada	Matriz de confusión		
.001	20		**0**	**1**
		0	**61**	27
		1	7	**23**
	15	**0**	**72**	4
		1	39	**3**
	10	**0**	**88**	1
		1	29	**0**
.01	20	**0**	**48**	36
		1	9	**25**
	15	**0**	**61**	24
		1	6	**27**
	10	**0**	**67**	23
		1	6	**22**
.05	20	**0**	**61**	21
		1	9	**27**
	15	**0**	**53**	28
		1	10	**27**
	10	**0**	**58**	31
		1	3	**26**

Results

Tasa de aprendizaje	Atributos de entrada	Exactitud	Precisión	Recall	F
.001	20	0.7119	0.4600	0.7667	0.5750
	15	0.6356	0.4286	0.0714	0.1224
	10	**0.7458**	0.0000	0.0000	**0.0000**
.01	20	0.6186	0.4098	0.7353	0.5263
	15	**0.7458**	0.5294	0.8182	**0.6429**
	10	**0.7542**	0.4889	0.7857	**0.6027**
.05	20	**0.7458**	0.5625	0.7500	**0.6429**
	15	0.6780	0.4909	0.7297	0.5870
	10	0.7119	0.4561	0.8966	0.6047

Conclusion

The most efficient model was the one with 15 variables with a learning rate of .01, with an effectiveness of 74.58% and an F index of .6429, which makes the model solid enough but will require the introduction of new variables seeking that the identification be more efficient so that it could help those interested in a more timely manner

Results

This knowledge model will be implemented in the Counseling System that the Institute has developed in previously to provide the professors and Counseling chief with an academic panorama expected of the students and to be able to make the most efficient and punctual monitoring process of those students that are more likely to drop out school.

References

Al-Twijri, M., & Noamanb, A. (2015). A New Data Mining Model Adopted for Higher Institutions. *Procedia Computer Science*(65), 836 – 844.

Luan, J. (2002). Data mining and its applications in higher education. *Journal of New Directions for institutional Research*, 17-36.

Riquelme, J., Ruiz, R., & Gilbert, K. (2006). Minería de Datos: Conceptos y Tendencias. *Revista Iberoamericana de Inteligencia Artificial, 10*(10), 11-18.

ECORFAN®