

Fig. 2. Archivo de entrenamiento “train.csv”.

Id	MSSubC	MSZoning	LotFrontage	LotArea	Street	Alley	LotShape	LandContour	Utilities	LotConfig	LandSlope	Neighborhood	Condition1	
1461	20	RL	80	1822	Pave	NA	Reg	Lul	AIUPub	Inside	Gil	Nlmes	Feedr	
1462	20	RL	81	14207	Pave	NA	IR1	Lul	AIUPub	Corner	Gil	Nlmes	Norm	
1463	60	RL	74	13830	Pave	NA	IR1	Lul	AIUPub	Inside	Gil	Gibson	Norm	
1464	60	RL	78	9978	Pave	NA	IR1	Lul	AIUPub	Inside	Gil	Gibson	Norm	
1465	120	RL	43	5505	Pave	NA	IR1	HLS	Lul	AIUPub	Inside	Gil	StonVil	Norm
1466	60	RL	75	10000	Pave	NA	IR1	Lul	AIUPub	Corner	Gil	Gibson	Norm	
1467	20	RL	NA	7980	Pave	NA	IR1	Lul	AIUPub	Inside	Gil	Gibson	Norm	
1468	60	RL	63	8402	Pave	NA	IR1	Lul	AIUPub	Inside	Gil	Gibson	Norm	
1469	20	RL	85	10176	Pave	NA	Reg	Lul	AIUPub	Inside	Gil	Gibson	Norm	
1470	20	RL	70	8400	Pave	NA	Reg	Lul	AIUPub	Corner	Gil	Nlmes	Norm	
1471	120	RL	26	5858	Pave	NA	IR1	Lul	AIUPub	FR2	Gil	Nlmes	Norm	
1472	180	RM	21	1880	Pave	NA	Reg	Lul	AIUPub	Inside	Gil	Bldg	Norm	
1473	180	RM	21	1880	Pave	NA	Reg	Lul	AIUPub	Inside	Gil	Bldg	Norm	
1474	180	RL	24	2280	Pave	NA	Reg	Lul	AIUPub	FR2	Gil	NPVil	Norm	
1475	120	RL	24	2280	Pave	NA	Reg	Lul	AIUPub	FR2	Gil	NPVil	Norm	
1476	60	RL	100	12858	Pave	NA	IR1	Lul	AIUPub	Inside	Gil	Ndgrh	Norm	
1477	20	RL	54	12883	Pave	NA	IR1	Lul	AIUPub	Corner	Gil	Ndgrh	Norm	
1478	20	RL	90	15220	Pave	NA	Reg	Lul	AIUPub	Inside	Gil	Ndgrh	Profl	
1479	20	RL	79	14122	Pave	NA	IR1	Lul	AIUPub	Inside	Gil	Ndgrh	Norm	
1480	20	RL	110	14300	Pave	NA	Reg	HLS	AIUPub	Inside	Mod	Ndgrh	Norm	
1481	60	RL	85	13850	Pave	NA	Reg	Lul	AIUPub	Corner	Gil	Ndgrh	Norm	
1482	120	RL	41	7132	Pave	NA	IR1	Lul	AIUPub	Inside	Gil	Ndgrh	Norm	
1483	20	RL	100	18434	Pave	NA	IR1	Lul	AIUPub	Corner	Gil	Gibson	Norm	
1484	120	RL	43	3263	Pave	NA	Reg	Lul	AIUPub	Inside	Gil	Bldg	Norm	
1485	60	RL	67	13300	Pave	NA	IR1	Lul	AIUPub	Inside	Gil	Gibson	Norm	
1486	60	RL	63	8577	Pave	NA	IR1	Lul	AIUPub	Inside	Gil	Gibson	Norm	
1487	60	RL	60	17433	Pave	NA	IR2	Lul	AIUPub	CUSao	Gil	NoRidge	Norm	
1488	20	RL	73	8987	Pave	NA	Reg	Lul	AIUPub	Inside	Gil	Somest	Norm	
1489	20	FV	32	3215	Pave	NA	Reg	Lul	AIUPub	Inside	Gil	Somest	Norm	
1490	20	FV	84	10440	Pave	NA	Reg	Lul	AIUPub	Corner	Gil	Somest	Norm	
1491	60	RL	70	11820	Pave	NA	Reg	Lul	AIUPub	Inside	Gil	Somest	Norm	

Fig. 3. Archivo de prueba “test.csv”.

Ahora, utilizando la herramienta RapidMiner se procede a realizar la carga de los Dataset “train.csv” y “test.csv” utilizando el operador “Read CSV” como se muestra en la Fig. 4.

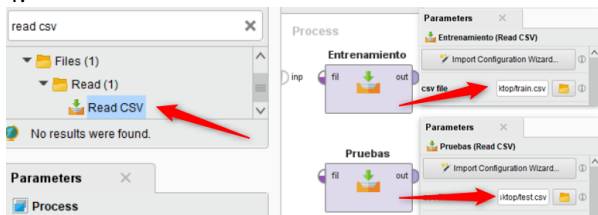


Fig. 4. Carga de Dataset “train.csv” y “test.csv”.

## B. Selección de los atributos

La selección de los atributos debe ser elegidos de acuerdo con la influencia que estos tienen en el atributo “SalePrice”. La Fig. 5 muestra los atributos para el Dataset entrenamiento y la Fig. 6 el Dataset prueba que se eligieron. En ambos casos se han elegido 7 atributos.

LotArea	YearBuilt	YearRemodA	GarageCars	GarageArea	FullBath	GrLivArea	SalePrice
8450	2003	2003	2	548	2	1710	208500
9600	1976	1976	2	460	2	1262	181500
11250	2001	2002	2	608	2	1786	223500
9550	1915	1970	3	642	1	1717	140000
14260	2000	2000	3	836	2	2198	250000
14115	1993	1995	2	480	1	1362	143000
10084	2004	2005	2	636	2	1694	307000
10382	1973	1973	2	484	2	2090	200000
6120	1931	1950	2	468	2	1774	129900
7420	1939	1950	1	205	1	1077	118000
11200	1965	1965	1	384	1	1040	129500
11924	2005	2006	3	736	3	2324	345000
12968	1962	1962	1	352	1	912	144000
10652	2006	2007	3	840	2	1494	279500
10920	1960	1960	1	352	1	1253	157000
6120	1929	2001	2	576	1	854	132000
11241	1970	1970	2	480	1	1004	149000
10791	1967	1967	2	516	2	1296	90000
13695	2004	2004	2	576	1	1114	159000
7560	1958	1965	1	294	1	1339	139000
14215	2005	2006	3	853	3	2376	325300
7449	1930	1950	1	280	1	1108	139400
9742	2002	2002	2	534	2	1795	230000

Fig. 5. Selección de datos de entrenamiento.

LotArea	YearBuilt	YearRemodA	GarageCars	GarageArea	FullBath	GrLivArea
11622	1961	1961	1	730	1	896
14267	1958	1958	1	312	1	1329
13830	1997	1998	2	482	2	1629
9978	1998	1998	2	470	2	1604
5005	1992	1992	2	506	2	1280
10000	1993	1994	2	440	2	1655
7980	1992	2007	2	420	2	1187
8402	1998	1998	2	393	2	1465
10176	1990	1990	2	506	1	1341
8400	1970	1970	2	525	1	882
5858	1999	1999	2	511	2	1337
1680	1971	1971	1	264	1	987
1680	1971	1971	1	320	1	1092
2280	1975	1975	2	440	2	1456
2280	1975	1975	1	308	1	836
12858	2009	2010	3	751	2	2334
12883	2009	2010	3	868	2	1544
11520	2005	2005	3	730	2	1698
14122	2005	2006	3	678	2	1822
14300	2003	2004	3	958	2	2696
13650	2002	2002	3	756	2	2250
7132	2006	2006	2	484	2	1370
18494	2005	2005	2	430	2	1324

Fig. 6. Selección de datos de prueba.

## C. Algoritmo Decision Tree

El algoritmo Decision Tree “es un árbol como una colección de nodos destinados a crear una decisión sobre la afiliación de valores a una clase o una estimación de un valor objetivo numérico. Cada nodo representa una regla de división para un atributo específico. Para la clasificación, esta regla separa los valores que pertenecen a diferentes clases, para la regresión los separa para reducir el error de manera óptima para el criterio del parámetro seleccionado.”[2].

La implementación del algoritmo y el árbol de decisión se muestra en la Fig. 7 y 8.

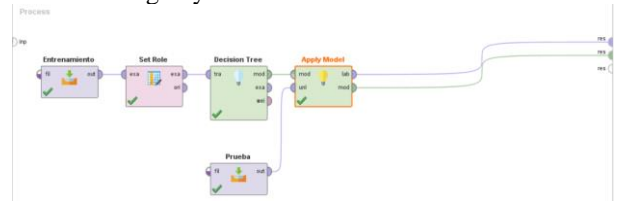


Fig. 7. Algoritmo “Decision Tree”.

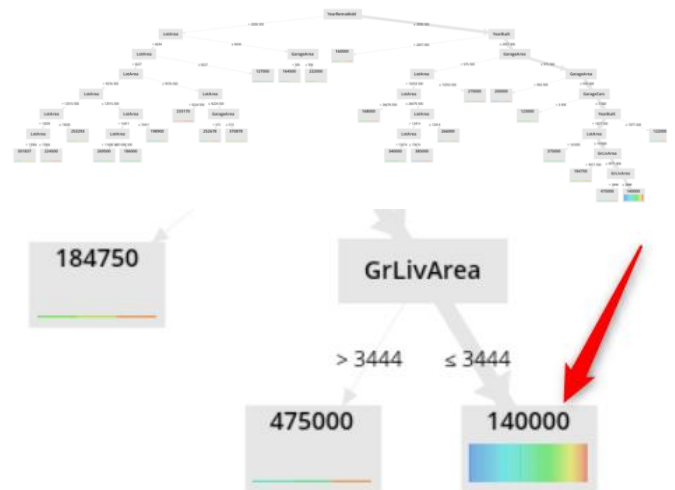
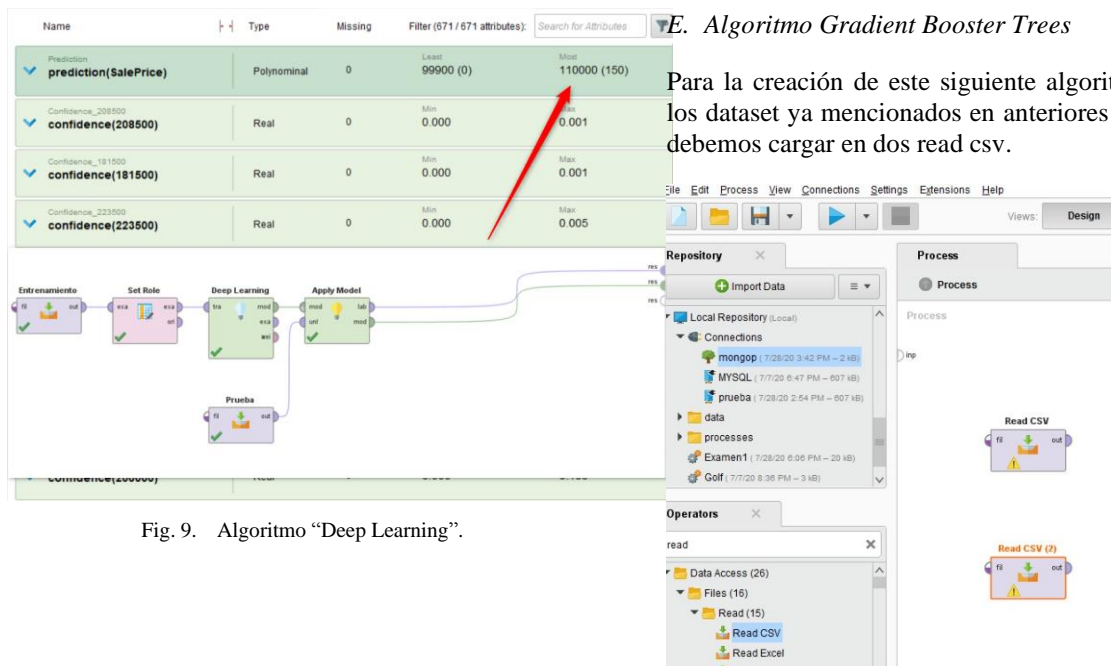


Fig. 8. Árbol de decisión.

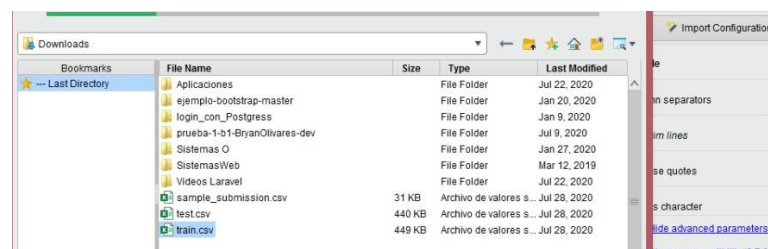
## D. Algoritmo Deep Learning

“Deep Learning (DL) que tiene una desviación estándar de +-0.5% (el segundo mejor por este indicador) con un Error Relativo de 9.9% (El segundo mejor por este indicador).”[2].

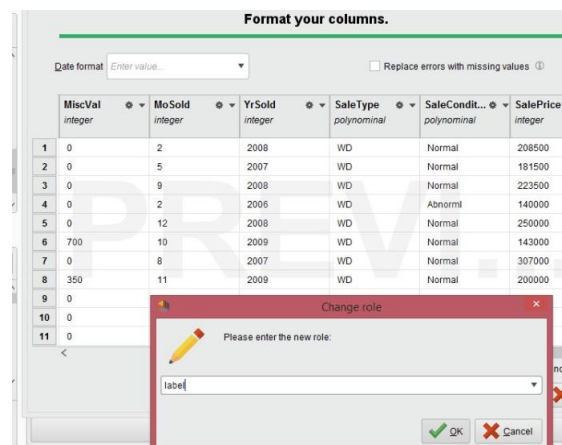
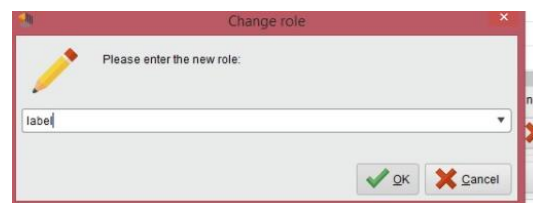
La Fig. 9 muestra el resultado de este algoritmo.



Para cargar los data set nos vamos al apartado de Import Configuration Wizard e importamos primero a train



Al momento de cargar el train debemos cambiar el rol a la columna price a label.



Para cargar el data set de test no va ser necesario cambiar el rol ya que es la columna que vamos a predecir.

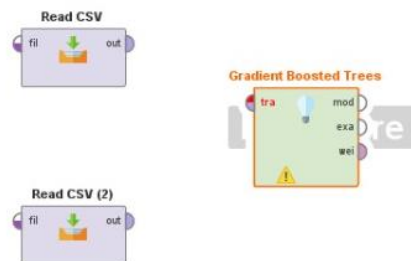
**Format your columns.**

Date format:  ☒ Replace errors with missing values ⓘ

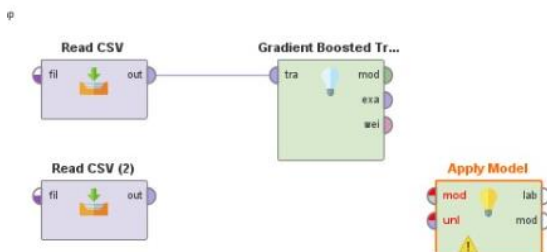
	MiscFeature	MiscVal	MoSold	YrSold	SaleType	SaleCondit...
	polynomial	integer	integer	integer	polynomial	polynomial
1	NA	0	6	2010	WD	Normal
2	Gar2	12500	6	2010	WD	Normal
3	NA	0	3	2010	WD	Normal
4	NA	0	6	2010	WD	Normal
5	NA	0	1	2010	WD	Normal
6	NA	0	4	2010	WD	Normal
7	Shed	500	3	2010	WD	Normal
8	NA	0	5	2010	WD	Normal
9	NA	0	2	2010	WD	Normal
10	NA	0	4	2010	WD	Normal
11	NA	0	6	2010	WD	Normal

< >

En este momento vamos a seleccionar el algoritmo en este caso gradient el que tiene menos porcentaje de error.



Para aplicar este modelo gradient vamos a unirlo con Apply Model.



Ahora si todo el procedimiento se completó correctamente debemos correrle y nos aparecerá el siguiente cuadro con los resultados.

Row No.	prediction(S...	Id	MSSubClass	M
1	160805.956	1461	20	R
2	164880.996	1462	20	R
3	183405.202	1463	60	R
4	182942.268	1464	60	R
5	198206.571	1465	120	R
6	179955.696	1466	60	R
7	172108.024	1467	20	R
8	181311.247	1468	60	R
9	180646.298	1469	20	R
10	155572.697	1470	20	R
11	179050.649	1471	120	R
12	149789.472	1472	160	R
13	155554.527	1473	160	R
14	168046.768	1474	160	R

Ahora vamos a la creacion de las conecciones para poder tener los datos en mongo db y en mysql.

**Edit connection - prueba (No responde)**

Info Setup Advanced Driver Sources

Database system: MySQL ⓘ

User: root ⓘ

Password: start typing ⓘ

☒ Configure URL automatically

Host: localhost ⓘ

Port: 3306 ⓘ

Database: prueba ⓘ

URL: jdbc:mysql://localhost:3306/prueba

An injected parameter is a parameter whose value is provided by an external source.

### III. CONCLUSIONES

El algoritmo de “Decision Tree” nos da una predicción del precio de venta es 140000 aproximadamente y en el algoritmo “Deep Learning” es de 110000 el mayor precio de venta.

### REFERENCIAS

- [1] “Minería de datos: Modelos predictivos con RapidMiner.” <http://www.tradingsys.org/mineria-de-datos-modelos-predictivos-con-rapidminer> (accessed Jul. 28, 2020).
- [2] M. René and P. Eddy, “Minería de Datos Caso de Estudio House Prices.”