Results of RandomForest Modeling

Last amended: 28th Oct, 2019

My folders: D:\data\OneDrive\Documents\knime-workspace\final_for_class\randomForest D:\data\OneDrive\Documents\big_mart_sales_problem

https://datahack.analyticsvidhya.com/contest/practice-problem-big-mart-sales-iii/

- 1. bigmartSales_randomForest model-I
 - a. Fill in missing values
 - b. Discretize MRP

0 0 0 M M M M	0070 0	ooy 👐 ۷		2	BED HRHE														
	A 3 naurim1	A Welcome to	KNIME Analytic	Diatform	A O biomartSal	er, randomForest mod							A Da S						
E	A st poysini i	Welcome to	Province Analytic	, Figuronn	21 C. Diginaritari								A0 00 2						
~ B⁄		Column Filter	Missing Val	Discretize ue numeric c	multiple column£ounting L	oop Start Partition	F	andom Forest Predi (Regression)	ctor Numeric Scorer	Loop End	Row Filter	Single sampl t-test							
> inal_for_class ^		<mark>+</mark> ₩►	→ ?	→,	► → (ً ► → 🗄		- (* *)	■ ²¹ ★	-+ O		<mark>∧</mark> ≗	wish -						
> decision_tree													Title D						
> feature_selecti		Node 25	Node 26	MR	RP loop 20	00 times 80:20		Predict	RMSE	Node 22	Include rows by	row id Only of Predicted valu	Descript						
 randomForest higmartSal 	File Reader	Statistics					Random Forest Le	amer)					No desc						
▲ bigmartSal	₽	+ 4 -					+ *						Tags						
Results of I								, 					No tags						
📋 ~\$sults of	Node 1	Node 17					Trees: 500						Links						
metanodes													No links						
ok_group Y		Linear Correlation	n										Creation						
>													Creation						
ow Coach 🐹 😐 🗖													Author						
E)		Node 2																	
		(local)																	
lepository 🗆 🗆		+ 55																	
retize																			
		Node 5											۲.						
													A KN 2						
at the set for set	<																		
natches found.	< E Outline ⊠			-	° 🗆 📮 Cons	ole 🛛						> 🔍 🚮 🚱 🛛							
natches found.	<	****			Cons	ole 🛿						> 🖹 🛃 🕅 🕅	,						
natches found.	<	• • • •		-	Cons KNIME C :a1) 0:	ole 🛛 Console 5 Some col	umns are ignored	: too many/missi	ng nominal value	5.		> R I P [
matches found.	< BE Outline X	••••	о #	- -	Cons KNIME C al) 0: c	ole 22 console 5 Some col	umns are ignored	: too many/missi	ng nominal value	s.	🛦 ^ ¥ 4) (m) (m) (m) (m) (m) (m) (m) (m) (m) (m	: 13:16 G 28-10-21						
natches found. P Type here to sea Je Sample T riptive Statistic:	e BE Outline 23		О Н		° □ Cons KNIME (:al) 0: <	ole 23 ionsole 5 Some col	umns are ignored	: too many/missi	ng nominal value	s. A Ŷ	▲ ^ ♥ «		i G 13:16 28-10-20						
natches found.	e BE Outline 23	N	O ⊨i Missing (count	Construction Const	ole 23 ionsole 5 Some col 1 Co	ums are ignored	: too many/missi	ng nominal value	s.	<u>∧</u>	> R I R C	: G 13:16 G 28-10-20						
Natches found. Type here to sea gle Sample T riptive Statistic: ediction (Item_O	c BE Outline 23 rrch C-Test s uttlet_sales)	N 200	O H Missing (count 0	Construction Co	ole 23 insole 5 Some col 1 Standar 3	umns are ignored	: too many/missi	ng nominal value Image: model	s. 🔺 😵 .	<u>∧</u>) (k () () (() () ())(())))))))	: 13:16 326-10-24						
Anatches found. Descriptive Statistics riptive Statistics ediction (Item_Or le Sample Test	c BE Outline 23 If the state of the stat	N 200	O Hi Missing (count 0	Construction of the second sec	ole 23 onsole 5 Some col 20 Standar 3	umns are ignored	: too many/missi	ng nominal value	s. 🛕 💱 _	<u>∧</u>		G 13:16 C 28-10-20						
Anatches found. Type here to sea gle Sample T riptive Statistic: ediction (Item_O e Sample Test dence Interval (C	e utilet_sales)	N 200 ty: 95.0%	O ⊨i Missing (Count 0	* Construction Con	ole 23 ionsole 5 Some col 20 20 20 20 20 20 20 20 20 20 20 20 20	ums are ignored	: too many/missi	ng nominal value	s. A 🌮	<u>∧</u>		G 13166 28-10-X						
natches found. O Type here to sea gle Sample T riptive Statistic: ediction (Item_O le Sample Test idence Interval (C	e BE Outline 23 Inch T-Test s uttlet_sales)	N 200 ty: 95.0% Test Va	O Hi Missing (count 0	Construction Co	ole 23 onsole 5 Some col 2 Standar 3 Standar	umns are ignored CR () d Deviation 8.752 ailed) Me	: too many/missi	ng nominal value The second	s. ▲ 😵	▲ ^ ♥ @ d) CI (S a s a c	6 3 1316 6 28-10-20						
Anatches found. D Type here to sea gle Sample T riptive Statistics rightive Statis rightive Statistics rightive Statistics right	CI) Probabili utlet_sales)	N 200 ty: 95.0% Test Va	O H Missing (ulue 0.0 1,	Count 0 t 881.1495	Construction Co	ole 23 osole 5 Some col 2 Standar 3 p-value (2-t	ums are ignored d Deviation 8.752 ailed) Me 0.0	: too many/missi	ng nominal value	•. • • • • • • • • • • • • • • • • • • •	<mark>▲ へ ♥ ⊄</mark> d) CI (>	d) 76						

Figure 1: Note that this result gives the least std error of mean even though Mean is highest.

2. bigmartSales_randomForest model-II

- a. Discretize MRP
- b. Missing values filled in for 80% and use the model for rest 20%
- c. Normalize outputsales using sqrt()

	🛕 3: paysim1	Welcome to I	KNIME Analytics	Platform 🔥 🔥 🗛	bigmartSales_randomF	Forest model-II	×							
<u>.</u>														
8 10														
9		Statistics			м	lissing Value (Apply)								
2		+ N E				7								
2														
		Node 1/	scretize multipl	le l		Node 13 Rand	tom Forest L Rai	nim Forest Predi	ctor				Single sample	
	File Reader	Math Formula n	umeric column	Counting Loop Start	t Partitioning M	lissing Value	(Regression)	(Regression)	Math Formula	Numeric Scorer	Loop End	Row Filter	t-test	
	B • •	▶ <mark>®</mark> ►	,	• I •		•?	**	ו••	► <mark>180 ►</mark>	• °**	• • •	► <mark>⇒</mark> ►	. N ^e ∫	
	Node 1	sqrt(item_Outlet_Sale	s) item_MRP	loop 100 times	80:20	Node 6	Node 10	Node 15	Node 19	Node 16	Node 23	Node 24	Node 25	
		Try also: log(target) Math Formula	Statistics											
		→ fix →	+ 4 +											
		Node 20	Node 21											
														1
	, Стур	e here to search		0) H ᇘ I	🤨 💽 💌	1 🔶 🗢	₩ - (R 🚯 🤞	4 🤪 🎵		 \$ 	へ 芋 🍖 🖮 🗔 🕪 🗈	13:17
	,О тур	e here to search		0) HI 📻 I	1 💉 💌	1 🔶 🖻	1 9¥ 🔲 (R 🛛 👈	4 🦻 🎵		\$	스 북 💊 🖮 🗔 다) 타	13:17 NG 28-10-20
4	, С _{Тур}	e here to search	iissiiiy va	0) = =	1 🛛	1 🔶 🖻) 🦋 🗖 (R 🖲 🧕	a 🦻 🎵		\$	^ 葉 @ 毎 見 ゆ) ₽	13:17 NG 28-10-20
4	, Отур	e here to search	ussiiy va (Apply)	0) 🖽 📻 (1	1 🔶 🖍	I 🧩 🗖 <	R 🚯 🧕	4 🤤 🎵		\$	◆ 葉 4 ● 〒 4 ● ●	13:17 NG 28-10-20
Test	€ Type	e here to search s - 5:25 - Sing	(Apply) Ie sample	o t-test) # 👼 (1 🛛 🖷	1 🔶 🗅	I 🛪 🗖 (R 🖲 🥌	4 2 7		\$	· ♥ @ @ 炅 0) B	13:17 NG 28-10-20
Test	P Typ statistic	e here to search s - 5:25 - Sing	(Apply) (Apply) le sample	O t-test) # 📷	1 🛛 🖷	1 🔶 🖻	1 % 🗖 (R 🖪 🤞	4 3 7		\$	◇ 葉 義 陽 巨 에 B — □	13:17 vG 28-10-20
Test		s - 5:25 - Sing	(Apply) (Apply) le sample	o t-test		0 🛛	1	96 🦳 🤇	R 🚯 🤲	4 2 7		\$		13:17 NG 28-10-20
Test	C Type statistic Sample	s - 5:25 - Sing	(Apply) (Apply) le sample	t-test			1 🔶 ፍ	96	R 🖲 🤲	4 2				13:17 28-10-24
Test	> Type statistics Sample btive Stat	re here to search s - 5:25 - Sing T-Test istics	(Apply) (Apply) le sample	t-test		1 🛛		1 % - (R 🖲 🤲	4 2				13:17 NG 28-10-20
Test	P Type statistics Sample	s - 5:25 - Sing T-Test istics	iissiily va (Apply) le sample	t-test		0 🛛	∎ ♠ 🗲	₩ 	R	4 ÷				* *G 13:17 28-10-20 X
Test	Sample	s - 5:25 - Sing T-Test istics	(Apply) (Apply) le sample	t-test	E HI E	0 🖭 💌] ∲ ≤	Standard D	R 🕑 🤙	Standard	Error Mean			* NG 13:17 28-10-20
Test	Sample	s - 5:25 - Sing T-Test istics	(Apply) (Apply) le sample	t-test	Esing Count	Mean 1.131	6415	standard D	R 🕑 🍏	J 2 1.	Error Mean			** NG 13:17 28-10-24
Test gle scrip	> Type statistics Sample ptive Stat	s - 5:25 - Sing T-Test istics	(Apply) (Apply) le sample	N Mis 200	sing Count	Mean 1,131.	6415	Standard D	eviation 28.1243	▲ ② 1. Standard	Error Mean 1.9887			13:17 NG 28-10-24
Test gle scrip	> Type statistics Sample prive Stat	s - 5:25 - Sing T-Test istics	II-SSIIIY VA (Apply) le sample	N Mis 200	sing Count	Mean 1,131.0	6415	Standard D	R 🕑 🍏	▲ ② 1. Standard	Error Mean 1.9887			13:17 NG 28-10-24
Test gle scrip gle	> Type statistics Sample ptive State liction(Ite Sample	s - 5:25 - Sing T-Test istics m_Outlet_Sa Test	(Apply) (Apply) le sample	t-test N Mis 200	Hi a	Mean 1,131.0	6415	Standard D	R 🕑 🧶	Standard	Error Mean 1.9887			13:17 28-10-24 X
Test gle crij red	> Type statistics Sample ptive State liction(Ite Sample ence Inte	s - 5:25 - Sing T-Test istics m_Outlet_Sa Test rval (CI) Prot	(Apply) le sample ales)	N Mis 200	Hi mi	Mean 1,131.0	6415	Standard D	R 🕑 🧶	Standard	Error Mean 1.9887			13:17 28-10-24 X
Test gle crij red	Sample diction(Ite Sample ntere Inte	e here to search s - 5:25 - Sing T-Test istics m_Outlet_Sa Test irval (CI) Prot	(Apply) le sample ales)	N Mis 200 .0%	Hi and state of the second sec	Mean 1,131.0	6415	Standard D	eviation 28.1243	Standard	Error Mean 1.9887			xG 13:17 28-10-24 X
Test gle scrip gle	C Type c statistic Sample otive Statistic Sample inction(Itte Sample ence Inte	e here to search s - 5:25 - Sing T-Test istics m_Outlet_Sa Test rval (CI) Prot	(Apply) le sample	N Mis 200 .0% Test Value	sing Count 0	Mean 1,131.0	6415	Standard D	R Image: Second se	Standard	Error Mean 1.9887	er Bound)		: 1317 2810-22 X

Figure 2: An improvement over result of !st model. Note the limits of confidence interval. Its limits are less than Model 1. It is the best model.

- 3. bigmartSales_randomForest model-III
 - a. Generate missing values using predictive analytics
 - b. Also normalize outputsales and discretize MRP

	Welcome to KNIME Analytics Plat	form 🔥 0: bi	gmartSales_rand	lomForest model	-II 🔥 5: b	oigmartSal	es_randomForest_m	odel-III 🗙						
Mis File Reader	sing values imputation E with random Forest Math Formula)iscretize multiple numeric columns	Column Filter	r Counting Loop	Start Partit	Rand	form Forest Literation (Regression)	(Regression)	lictor Math Formula	Numeric Scorer	Loop End	Row Filter	Single sample t-test	
B +			► ±± ►	I	<u>ا</u>		+ 25		▶ 60 ►	• °!	+0+	→ <u></u>	Ala C	
Node 1	Takes time sqrt(Item_Outlet_Sal	es) MRP	Node 39	Node 3	1 Nod	de 12	Node 10	Node 15	Node 19	Node 16	Node 32	Node 33	Node 34	
		Statistics					\setminus /							
		· + ₩ È					\sim							
		Node 21												
														-
		_								1			1	12110
, О _{Туре}	here to search	0	Hi 📘	I 🚺 🛃	ف 🙍	Ş	n 🗖 👁	8 单	4 🤝 🗾	····	▲ 🕸 ▲ ^		… にの 28-	13:18 -10-2019
, О _{Туре}	here to search	0	H 📻	I 🚺 💽	•		¥ 🗖 🧟	8 🔌	🦉 🥪 😕		\$ \$ <u>\</u> ^	₩ 46 10	に こ 1 1 1 1 1 1 1 1 1 1 1 1 1	13:18 10-2019
C Type	here to search	O Node 12 mple t-test	Hi 📻	10.10. V	Node		Nod	8 🔌	A Sode 16		▲ 😵 🔺 ^	₩ 🇠 🖘 Node	い い い の こ し	13:18 10-2019 Node X
✓ Type 30 Test sile	here to search Node 31 tatistics - 6:34 - Single sa	O Node 12 mple t-test	Hi 📻	1e 10 🗸	Node		Nod	8 👲	A 🍣 🔽 Node 16		▲ 😵 🛕 ^	- ₩ @ \= 	: () () () () () () () () () () () () () (13:18 10-2019 Node X
✓ Type 30 ∆ Test sile	here to search Node 21 tatistics - 6:34 - Single sa	O Mode 12 mple t-test	Hi 🗖	10.10 V	Node	▲ 15	Nod	8 🖕	Node 16		ada 32	t‡ a⊛ tan Node _	⊑ Ф) ENG ₂₈ . 23.	13:18 10-2019 Mode X
C Type	here to search Norte 21 I tatistics - 6:34 - Single sa ample T-Test	O Node 12 mple t-test	H Nor	1e 10	Node	15	Nod	8 🥹 e 19	Ma 🥏 📕		▲ 🖗 🛕 ^	te de m Node	:	13:18 10-2019 Nocte X
C Type 20 ▲ Test s ile Single Sa	here to search Norte 21 I tatistics - 6:34 - Single sa ample T-Test	O Note 12 mple t-test	H .	10 10 V	Node	15 15	Nod	8 6	Node 16		▲ ♥ 🛕 ^	të a⊚ ≌ Node	: () ENG 28- 22	13:18 10-2019 Norde X
✓ Type 20 20 Test state ile Single Sa Description	here to search Norte 21 I tatistics - 6:34 - Single sa ample T-Test we Statistics	O Node 12 mple t-test	Hi 🗖	10 10 V	Node	▲ 15	Nod	8 👲	Node 16		▲ ♥ 🛕 ^	♥ @ ₪ 	₩ 40 ENG 28	13:18 10-2019 Mode X
✓ Type 30 Test stile Single Sa	here to search Norte 21 I tatistics - 6:34 - Single sa ample T-Test we Statistics	O Node 12 mple t-test	Hi Nor	10 10 🖉	Node	• 15	Nod	A 10 A 10	Node 16	ard Error M	ode 32	₩ @ %) Node _	; ₽ Φ) ENG ₂₈ ; 22	13:18 10-2019 Node X
C Type	here to search Norde 21 I tatistics - 6:34 - Single sai ample T-Test we Statistics dion(Item_Outlet_Sales)	Node 12 mple t-test	Hi Nor Missing	1 1 1 ×	Mean	■ 15	Nod	(1) (2) (6) (3) (4)	Node 16	ard Error M	ean 835	₩ @ @ Nodo –		13:18 10-2019 Node X
	here to search	O Mode 12 mple t-test N 200	Hi Nor	1 (1) 10 10 / 1 Count 0	Mean 1,146.1	15 1965	Nod	(1) (2)	Node 16	ard Error M	ean 835	♥ @ %) Node _		13:18 10-2019 Nocto X
P Type 30 ▲ Test s ile Single Sa Descripti predic	here to search	O Node 12 mple t-test	H Nor	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Mean 1,146.1	15	Nod	(1) (2)	Node 16	ard Error M	ean 835	♥ @ %	і Г. Ф. вко ₂₈ 23.	13:18 10-2019 Mode X
Type Type Test s ile Single Sa predic Single Sa	here to search	0 Note 12 mple t-test N 200	H Nor	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	Mean 1,146.1	 ▲ 15 1965 	Nod	 19 10 10 10 26.63 	Node 16	ard Error M 1.8	ean 835	tt ang tao	:	13:18 10-2019 Node X
P Type 20 ▲ Test s ile Single Sa Descripti predic Single Sa Confiden	here to search	0 Note 12 mple t-test N 200 by: 95.0%	El Nor	1 1 2 2 10 10 7	Mean 1,146.1	965	Nod	 10 10 10 26.63 	Node 16	ard Error M	ean 835	tt ang tao	:	13:18 10-2019 Node X
P Type 20 Test s ile Single Sa Descripti predic Single Sa Confiden	here to search Node 31 tatistics - 6:34 - Single sai ample T-Test ve Statistics tion(Item_Outlet_Sales) ample Test ice Interval (CI) Probabili	0 Note 12 mple t-test N 200 by: 95.0% Test V	H Nor Missing	Count 0	Mean 1,146.1	9965	Nod		Node 16 Node 16 n Stand 6	ard Error M 1.8	ean 835	₩ @ 100 	i no 25 22	13:18 10-2019 Nocto X

Figure 3: This model falls in between Model 1 and model 2
