

Data Mining Milestone 5

WQD180036 TAN KAI XUAN

Training and Validation set

Partition Summary

Type	Data Set	Number of Observations
DATA	EMUS2.Ids4_DATA	205
TRAIN	EMUS2.Part3_TRAIN	102
VALIDATE	EMUS2.Part3_VALIDATE	103

Summary Statistics for Class Targets

Data=DATA

Variable	Numeric Value	Formatted Value	Frequency Count	Percent	Label
PriceLabel	.	Decrease	91	44.3902	
PriceLabel	.	Increase	83	40.4878	
PriceLabel	.	No Change	31	15.1220	

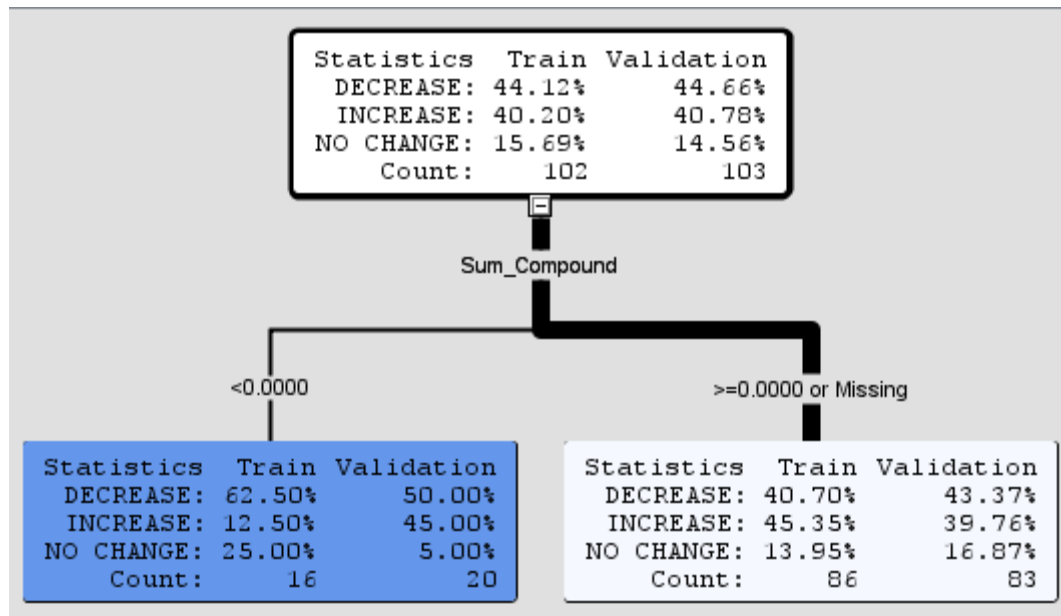
Data=TRAIN

Variable	Numeric Value	Formatted Value	Frequency Count	Percent	Label
PriceLabel	.	Decrease	45	44.1176	
PriceLabel	.	Increase	41	40.1961	
PriceLabel	.	No Change	16	15.6863	

Data=VALIDATE

Variable	Numeric Value	Formatted Value	Frequency Count	Percent	Label
PriceLabel	.	Decrease	46	44.6602	
PriceLabel	.	Increase	42	40.7767	
PriceLabel	.	No Change	15	14.5631	

Maximal Tree



Sum_Compound - Interval Split Rule

Target Variable: PriceLabel

Assign missing values to

☒ A specific branch 2

☐ A separate missing values branch

☐ All branches

Branches

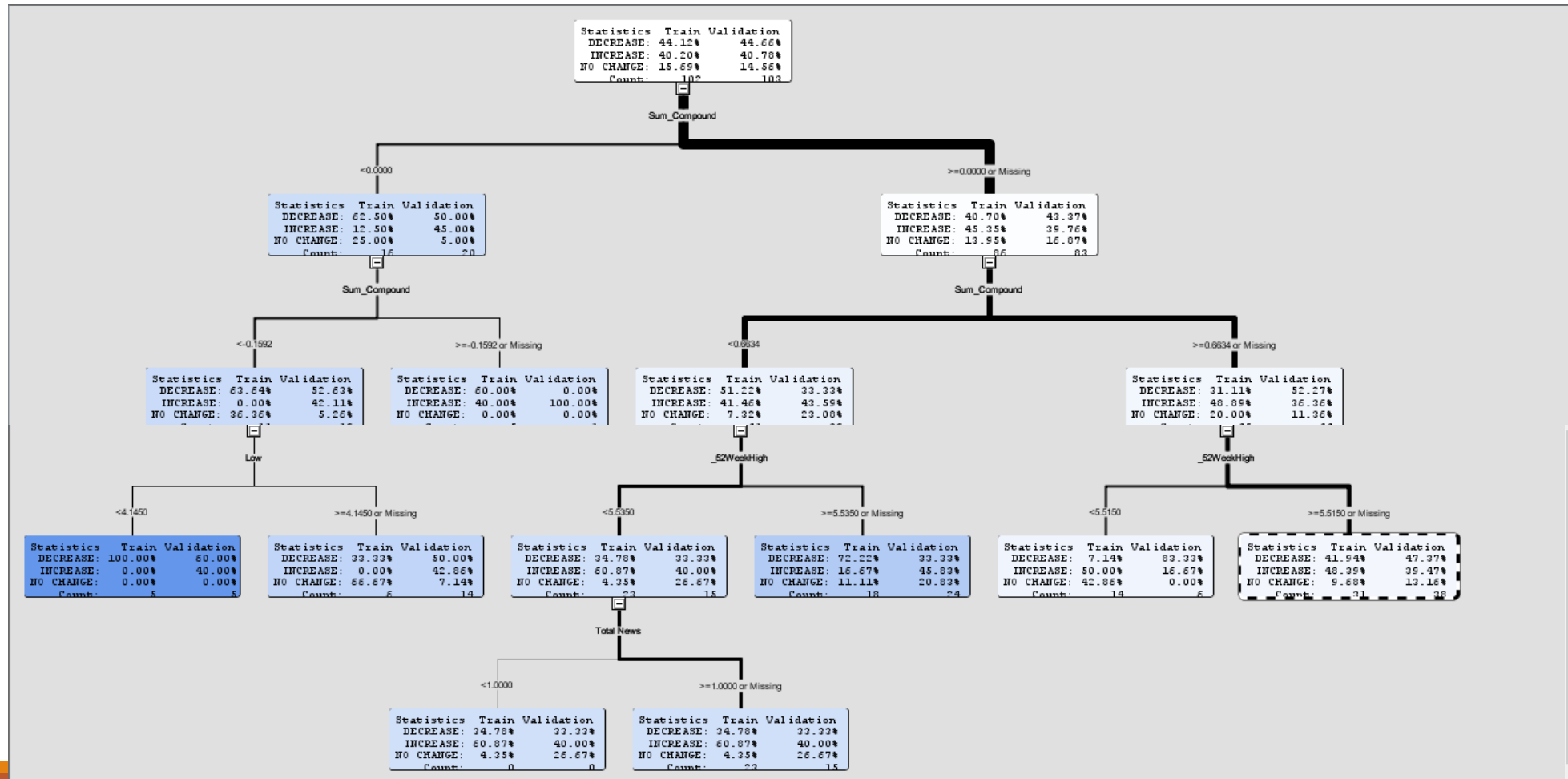
Branch		Split Point
1	<	0.0000
2	>=	0.0000

New split point:

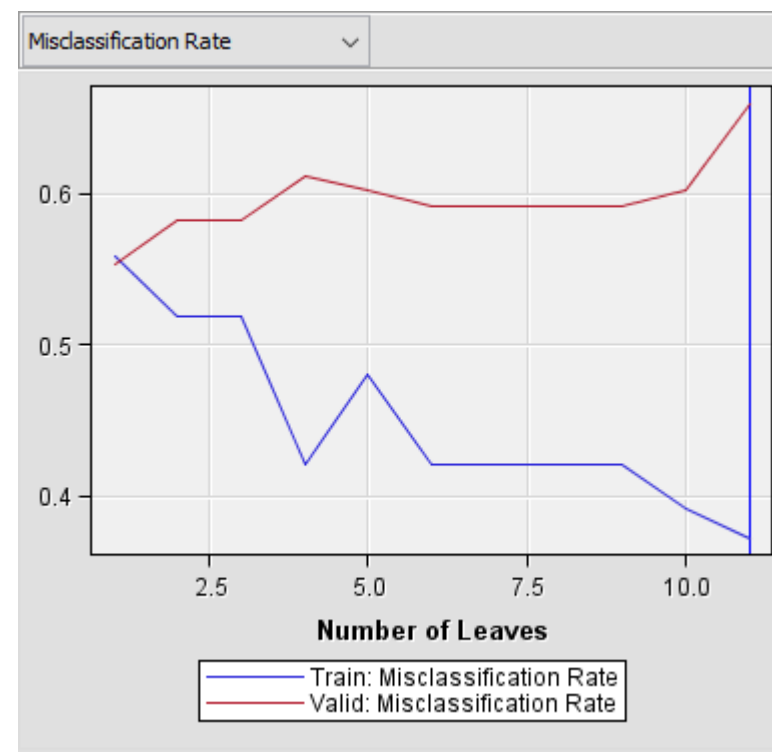
Add Branch Remove Branch

OK Cancel Apply Reset

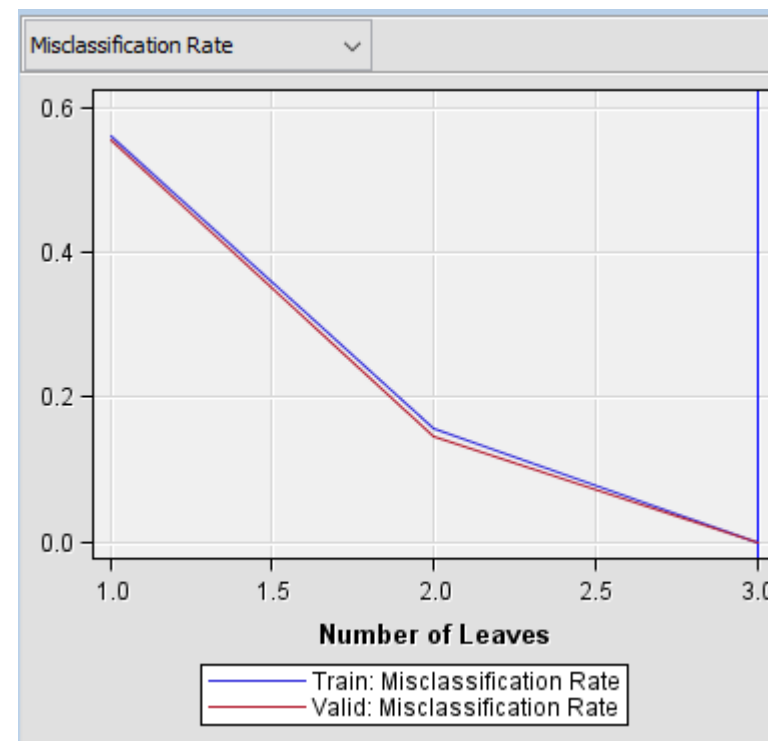
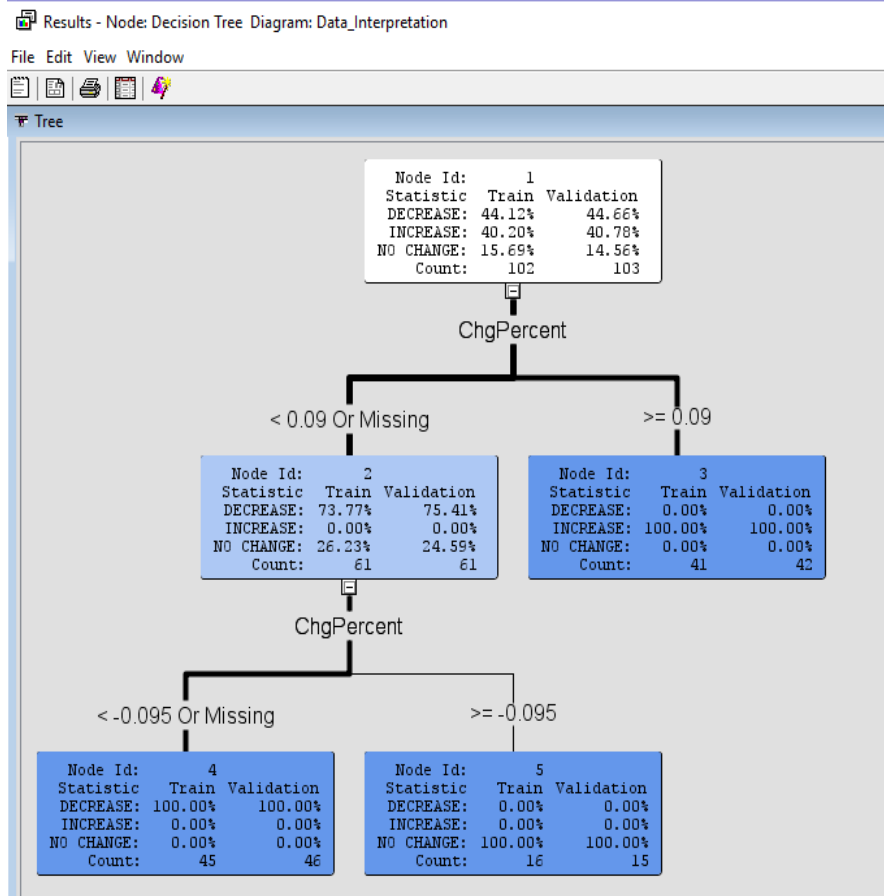
Maximal Tree



Maximal Tree



Decision Tree



Logistic Regression

Model Information

Training Data Set	WORK.EM_DMREG.VIEW
DMDB Catalog	WORK.REG3_DMDB
Target Variable	PriceLabel
Target Measurement Level	Nominal
Number of Target Categories	3
Error	MBernoulli
Link Function	Logit
Number of Model Parameters	28
Number of Observations	65

Type 3 Analysis of Effects

Effect	DF	Wald Chi-Square	Pr > ChiSq
Company	8	9.3282	0.3154
High	2	6.2058	0.0449
Last	2	11.5665	0.0031
Low	2	5.3587	0.0686
Open	2	4.7482	0.0931
Sum_Compound	2	3.1837	0.2036
Total_News	2	1.5514	0.4604
Volume	2	0.5350	0.7653
_52WeekHigh	2	2.8735	0.2377
_52WeekLow	0	0.0000	.

Logistic Regression

Fit Statistics						
Target	Target Label	Fit Statistics	Statistics Label	Train	Validation	Test
PriceLabel		_AIC_	Akaike's Inform...	209.1531	.	.
PriceLabel		_ASE_	Average Squar...	0.151912	0.207154	.
PriceLabel		_AVERR_	Average Error ...	0.513572	0.748154	.
PriceLabel		_DFE_	Degrees of Fre...	178	.	.
PriceLabel		_DFM_	Model Degree...	26	.	.
PriceLabel		_DFT_	Total Degrees ...	204	.	.
PriceLabel		_DIV_	Divisor for ASE	306	309	.
PriceLabel		_ERR_	Error Function	157.1531	231.1795	.
PriceLabel		_FPE_	Final Predictio...	0.196291	.	.
PriceLabel		_MAX_	Maximum Abs...	0.982999	0.99809	.
PriceLabel		_MSE_	Mean Square ...	0.174102	0.207154	.
PriceLabel		_NOBS_	Sum of Freque...	102	103	.
PriceLabel		_NW_	Number of Esti...	26	.	.
PriceLabel		_RASE_	Root Average ...	0.389759	0.455142	.
PriceLabel		_RFPE_	Root Final Pre...	0.443048	.	.
PriceLabel		_RMSE_	Root Mean Sq...	0.417255	0.455142	.
PriceLabel		_SBC_	Schwarz's Bay...	295.4242	.	.
PriceLabel		_SSE_	Sum of Square...	46.4852	64.01063	.
PriceLabel		_SUMW_	Sum of Case ...	306	309	.
PriceLabel		_MISC_	Misclassificati...	0.382353	0.475728	.

Logistic Regression - Stepwise

Step 0: Intercepts entered.

The DMREG Procedure

Newton-Raphson Ridge Optimization

Without Parameter Scaling

Parameter Estimates 2

Optimization Start

Active Constraints	0	Objective Function	66.388354237
Max Abs Gradient Element	4.440892E-15		

Optimization Results

Iterations	0	Function Calls	3
Hessian Calls	1	Active Constraints	0
Objective Function	66.388354237	Max Abs Gradient Element	4.440892E-15
Ridge	0	Actual Over Pred Change	0

Convergence criterion (ABSGCONV=0.00001) satisfied.

Likelihood Ratio Test for Global Null Hypothesis: BETA=0

-2 Log Likelihood		Likelihood Ratio		
Intercept Only	Intercept & Covariates	Chi-Square	DF	Pr > ChiSq
132.777	132.777	0.0000	0	.

Logistic Regression - Stepwise

Analysis of Maximum Likelihood Estimates

Parameter	Price Label	DF	Estimate	Standard Error	Wald Chi-Square	Pr > ChiSq	Standardized Estimate	Exp(Est)
Intercept	No Change	1	-1.0361	0.3510	8.72	0.0032		0.355
Intercept	Increase	1	-0.2985	0.2752	1.18	0.2781		0.742

Logistic Regression - Stepwise

Fit Statistics

Target=PriceLabel Target Label=' '

Fit Statistics	Statistics Label	Train	Validation
AIC	Akaike's Information Criterion	212.673	.
ASE	Average Squared Error	0.208	0.206
AVERR	Average Error Function	0.682	0.675
DFE	Degrees of Freedom for Error	202.000	.
DFM	Model Degrees of Freedom	2.000	.
DFT	Total Degrees of Freedom	204.000	.
DIV	Divisor for ASE	306.000	309.000
ERR	Error Function	208.673	208.679
FPE	Final Prediction Error	0.212	.
MAX	Maximum Absolute Error	0.831	0.831
MSE	Mean Square Error	0.210	0.206
NOBS	Sum of Frequencies	102.000	103.000
NW	Number of Estimate Weights	2.000	.
RASE	Root Average Sum of Squares	0.456	0.454
RFPE	Root Final Prediction Error	0.460	.
RMSE	Root Mean Squared Error	0.458	0.454
SBC	Schwarz's Bayesian Criterion	219.309	.
SSE	Sum of Squared Errors	63.539	63.597
SUMW	Sum of Case Weights Times Freq	306.000	309.000
MISC	Misclassification Rate	0.559	0.553

Event Classification Table

Data Role=TRAIN Target=PriceLabel Target Label=' '

False Negative	True Negative	False Positive	True Positive
16	86	0	0

Data Role=VALIDATE Target=PriceLabel Target Label=' '

False Negative	True Negative	False Positive	True Positive
15	88	0	0

Logistic Regression - Stepwise

