Soumya Joshi
Vandana Anand

MKT 568

Assignment 1, Spring 2020

Modules 1&2

Recommended Text: LaRose and LaRose (Chapters 1 and 5.6)

Due: February 2nd, 11:59 pm
Scoring: Indicated below in (_) (28 points total)

Question A.
a) (1) Classification
b) (1) Description
c) (1) Association
d) (1) Cluster Analysis
e) (1) Association
f) (1) Prediction

Question B.

a) Translating business narratives
b) Turning a narrative into a goal will make it so that we can turn it into a data mining problem
c) Select and Present the most important results

Question C.

a) MFG Month and MFG Year, because they both give the age of the vehicle which is already depicted in the column “Age in months”.
b) Automatic and Met_Color
c) For every one unit increase in KM, the price of a car will decrease by 0.015 units, keeping all other predictors constant. To a non-technical executive it can be explained that if the number of miles the car has been used goes up by 1, the price of the car will go down by 0.015 dollars. For example, for every 100 miles increase in the car, the price of the car will be less by 1.50 dollars.
d) Here is the coefficient table we obtained from running IBM Modeler. We obtained this table by setting all the variables, except Age, KM, and Fuel Type, to “none” as the question asks to use only the above listed variables as inputs:

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	20757.390	287.664		72.159	.000
	Age_08_04	-147.850	4.377	-.763	-33.783	.000
	KM	-.019	.002	-.200	-8.020	.000
	Fuel_Type_CNG	-600.884	711.302	-.016	-.845	.399
	Fuel_Type_Petrol	-585.856	240.593	-.054	-2.435	.015

We can estimate the effect of a car using diesel by finding that the amount of CNG is -600.884 less than the amount of diesel and the amount of petrol is -585.856 less than the amount of diesel. The higher priced car is the one using diesel because the petrol car is -585.865 less than the diesel car.
Also, if we calculate the predicted price for both types, we get,

Price of Diesel car = 20757.390(constant) – 147.850* Age – 0.019* KM + 0*(Fuel type = Diesel)

Taking arbitrary values of other variables we get,

Price of Diesel car = 20757.390 – 147.850(30) – 0.019(48000) + 0(1) = 15,409.89

Similarly ,

Price of Petrol car = 20757.390(constant) – 147.850* Age – 0.019* KM -585.856*(Fuel type = Petrol)

We have,

Price of Petrol car = 20757.390 – 147.850(30) – 0.019(48000) -585.856(1) = 14,824.034

Hence we can say that the price of Diesel Car will be more than that of Petrol one.

e) We have created three separate models each having age* a fuel_type as a derived variable. From the regression analysis, the car having the fuel type Diesel has the lowest effect on the car's price showing that the Diesel helps a car hold their value. Whereas for fuel type Petrol, the price of the car goes down by 77.3 dollars for every 1 year increase in age and for the fuel type CNG, the price of the car goes down by 35 dollars for every 1 year increase in age. Below are the coefficient tables we obtained by setting all the other variables to "none" and keeping Age*FuelType as inputs.

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	10714.528	134.903		79.424	.000
	Age*CNG	-35.097	24.217	-.055	-1.449	.148

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	14532.951	251.762		57.725	.000
	Age*Petrol	-77.322	4.534	-.543	-17.052	.000

Coefficients

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	10839.585	140.422		77.193	.000
	Age*Diesel	-25.571	7.787	-.123	-3.284	.001

f) Academic:

- The regression model has an R_2 near 80%.
- Both regression and tree models can be easily explained to our company's management and to car owners who dispute our appraisal.
- There are other models (e.g. bagged trees) whose predictions will be better than the regressions, in terms of mean squared error loss.
- A non-linear model will have less bias.

Executive:

- A tree model will not predict sales price very well.
- Three of the most important variables are Age_08_04, KM and Fuel Type
- The best model in this analysis gives a car appraiser a quick way of predicting the sale price of an individual, using no more than a calculator.