Data 640 – Spring 2019

Assignment 2

Support Vector Machine Analysis of Telco Customer Churn

Kenneth Lulie

University of Maryland University College

Professor Knode

March 18, 2019

**Support Vector Analysis of Churn using the Telco Dataset**

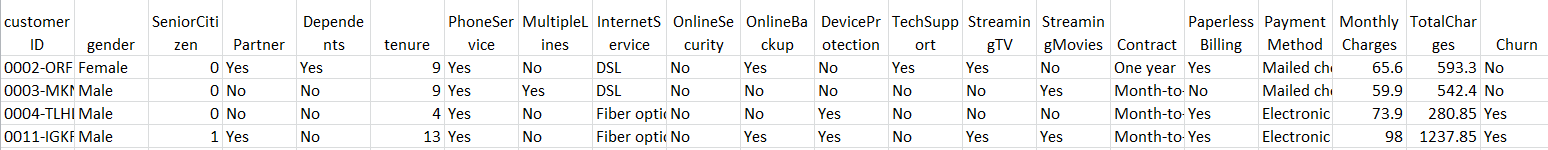
**Introduction**

A classic application of data science is to identify when a customer will stop using your service or product, also called churn. This prediction is valuable as it enables cost effective decisions to intervene to prevent that churn increasing the value of churn interventions (Han, 2011). As identifying churn is one of the most frequent uses of data science, newer techniques should be evaluated to see how effective the new model technique is for this purpose. Support Vector Machines or SVMs is a new technique that is beginning to be more widely used. The purpose of this analysis will be to simulate the usage of SVMs on customer churn to test the fitness of SVMs for identifying customer churn.

The first metric used for this analysis will be to compare the total misclassification rate of four SVM models. However, as churn is typically an unbalanced dataset another method should be used to compare its performance than mere misclassification. The second metric used will be to compare the results of a mock intervention analysis where a correct identification of a churning customer, or True Positive, is worth $4.00 to simulate the average expected value of that knowledge being used to prevent a churn while an incorrect identification of churn, or false positive, is worth -$1.00 to reflect the cost of an unsuccessful intervention. True Negatives and False Negatives are not rewarded or penalized. Additionally a decision tree model will also be created and scored to give a comparison of SVMs to existing techniques.

The data used in this analysis will be from the ‘Telco Churn Dataset’ provided by IBM. (IBM, 2015). This dataset consists of 6,893 rows with one unique identification variable, 18 independent variables representing different characteristics of a customer such as their gender, if they are a senior citizen, if they have a partner, or if they are a dependent as well as characteristics of their contract with Telco such as the duration of their tenure as a customer in months, which of the services provided by Telco are being provided, their payment method, contract type, monthly and total charges. A target binary variable labeled Churn is also provided with 5024 ‘No’ values representing 72.8% of the dataset and 1869 ‘Yes’ values representing 27.2% of the dataset **(Table 1)**.

Table 1. Sample rows of Telco Churn Dataset



**Telco Dataset Data Loading and Cleansing**

The data used in this analysis was downloaded from the IBM website in .csv format, and then imported into SAS Enterprise Miner using the File Import node. A complete picture of the SAS Enterprise Miner diagram used in this analysis can be seen in **Figure 1.** As the dataset was under 60,000 rows no sampling was necessary, and the entire dataset was loaded in. Each of the variables were reviewed using the explore function in SAS Enterprise Miner as well as manually in Microsoft Excel looking for data oddities, errors, or missing values **(Table 2).**

This review found that .159% of the total rows (11 observations) were missing values for the “TotalCharges” variable. On further review in excel it was noted that each of these rows had a ‘tenure’ value of 0 indicating that these customers had not yet been customers for a full month. This showed that these rows were likely simply customers who had not yet been charged for a month of service. As SVMs ignore observations with missing values, the impute node was used to convert these missing values to ‘0’.

The dataset was loaded using the Input Data node, and 1 variable was automatically rejected for having over 20 levels, which was the ‘customerID’ variable. The ‘Churn’ variable was manually set to the Role of target. The remaining variables were correctly classified as either Nominal, Interval or Binary.

**Telco Dataset Data Exploration and Preparation**

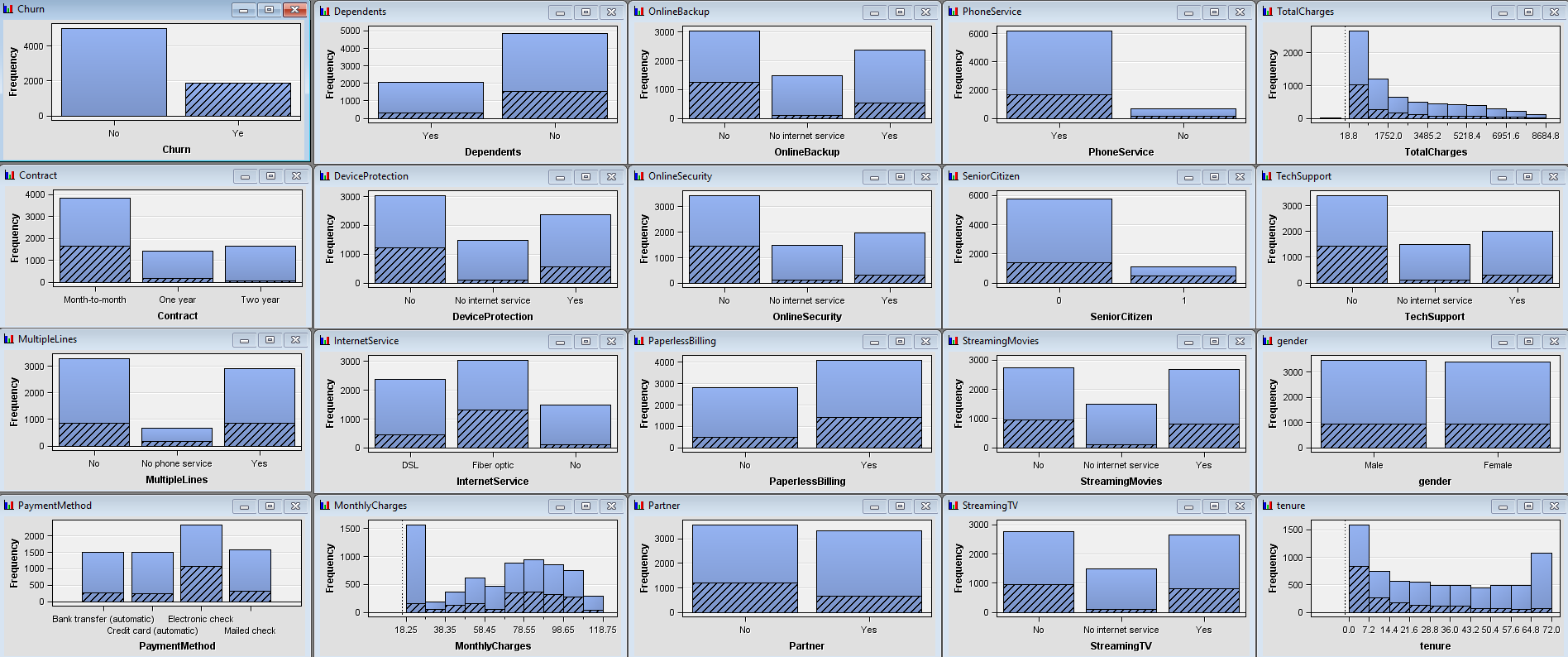
The data was reviewed using the explore function with the observations where the customer did churn shaded in the graphs showing the distribution of the variables **(Figure 2)**. As previously mentioned the dataset is skewed with approximately one out of four customers identified as having churned. This imbalance is addressed by adding a cost function evaluation as previously described to avoid the model being rewarded for simply predicting the most frequent category. The graphs appear to show some indications that month to month contracts, and contracts under 7 months old are the most likely to churn. Additionally, customers who have fiber optic service, no online backup, no OnlineSecurity package, and customers who pay by electronic check are also more likely to churn than the average. Further, it also appears that monthly charges higher than around $60 also appear more likely to churn. Finally, customers who are dependents were much less likely to churn than customers who are not dependents.

Figure 2. Data Exploration of Telco Datset

There are some common sense explanations for these observations. Higher monthly charges mean more incentive to save money by changing providers. Month to month contract holders have lower transaction charges to change providers. Further, customers without additional services may find it easier to change as well. The skewness of the variables were also reviewed and it was noted all skewness values were between -2 and 2 showing no transformations were necessary to reduce skewness **(Table 2)**.

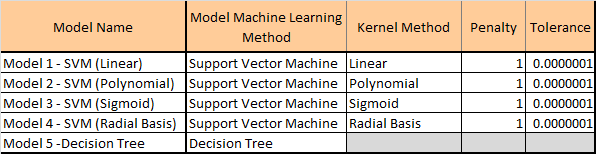
While SVMs require each input to be in a numerical form, no transforms for categorical variables is needed as SAS Enterprise Miner will automatically convert the categorical variables into dummy binary variables(SAS, 2018). Additionally, as SVMs are well known for being able to handle large amounts of variables no variable selection was used (Munnangi, 2015).

The data was then partitioned using the Data Partition node, with 70% or 4,823 observations going into the train set, 20% or 1,379 going into the test set, and 10% or 691 going into the validate set. It is important to separate data into these sets to avoid bias and over-fitting in your model scoring and selection (Han, 2011). SAS Enterprise Miner will train on the test data to build models, and then use the validation data set to compare them against each other. Then the test data set can be used to get a score of its unbiased performance on new data.

**SVM Models Developed**

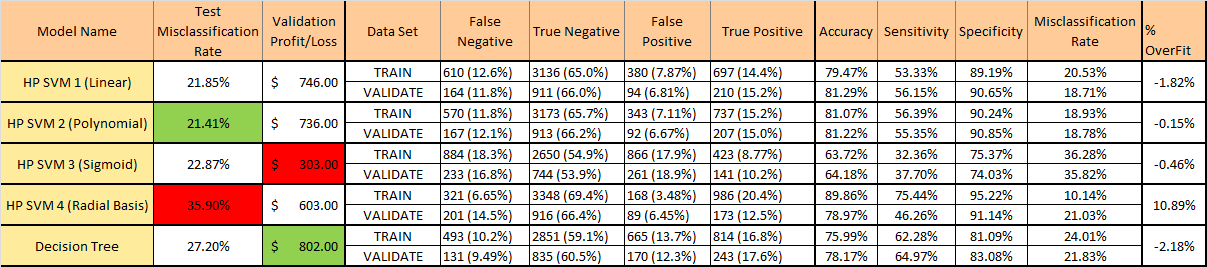
A Support Vector Machine is a classification Machine Learning data technique that operates by creating a linear decision surface, or hyperlane as a boundary to classify the data (Munnangi, 2015). The support vectors refer to the observations closed to the hyperlane. It is a supervised learning technique that requires labeled data to train off of, and is best at performing binary classification. It is also important to note that the input variables must be binary, ordinal, nominal, or interval but SAS Enterprise Miner will automatically transform categorical variables into dummy binary variables for this analysis (SAS, 2018). Additionally, observations with missing values are ignored during model training.

Among the advantages of using SVMs is that they do well with large numbers of variables and small samples and can be highly accurate. However, training time can be prohibitively large for very large datasets. Noisy data can be a problem for SVMs, but they can be tuned to adjust for this by adjusting the ‘margin’ or how much separation should be allowed when constructing the hyperplanes. A softer margin may allow for more misclassifications, but help avoid over-fitting (Han, 2011). The other parameter you can tune is using a kernel to transform the linear boundary into another dimension increasing the feature space and potentially allowing more powerful classification.

There were five different models developed for this analysis **(See Figure 3)**. Four of these models represented each of the Kernel methods available in SAS Enterprise Miner using the default parameters with the High Performance Support Vector (HP SVM) nodes, and the fifth was a Decision Tree also made by SAS Enterprise Miner with default settings using the Decision Tree Node. The SVM models differ only in the Kernel methods used to increase the feature space.

**Results – Polynomial Kernel SVM has best misclassification Rate, Decision Tree Best Profit**

The models were evaluated using the Model Comparison Node which selected from the models based on the lowest misclassification rate from the test data set **(Table 3)**. Additionally, the profit or loss was calculated for the validation data set. Based on this review, we can see that Model 2, the SVM with Polynomial kernel had the lowest misclassification rate on the test dataset, but that the Decision Tree had the highest computed profit for the validation dataset.

Table 3. Model Comparisons Results

Additionally, we can see that there was generally minimal over-fitting when comparing the training dataset to the test dataset with the exception of Model 4 – SVM with Radial Basis Kernel which had a powerful 10.14% misclassification rate on the training dataset, but 21.03% for the validation set and 35.9% for the test data set. It is worth noting that merely guessing no on every observation would yield only a 24% misclassification rate an improvement of 11.9% over Model 4’s performance on the test data. Additionally, Model 3 – SVM with Sigmoid Kernel had a poor misclassification rate on training and validation of 36.28% and 35.82% but surprisingly only a 22.87% on the test data set. We can also see from the Accuracy Sensitivity and Specificity that the SVM models had generally higher specificity than the decision tree model, but lower sensitivity. In other words the SVM models were better at only picking true positives, but they didn’t pick as many correct true positives.

From the profit computed we can see that the Decision Tree model won by $56 compared to the next runner up of Model 1 – SVM with Linear Kernel. The decision tree had more false negatives than the other models, but also the most true positives which is why it was the highest scoring model for this metric. However, a review of the Cumulative Lift of the models shows the SVMs may outperform the decision tree model when computing expected profit if only the first 20% of the dataset was to be used **(Figure 4).**

**Conclusions and Takeaways**

From the models produced, it would be hard to argue that the decision tree would not be the superior model. The decision tree had the highest profit, would be easier to implement, and would be easier to explain. However, if limiting the data to the 20% of the dataset containing the customers most likely to churn a SVM may perform better as shown by the Cumulative Lift comparison. Therefore SVMs, may have potential but should be compared to the performance of other techniques depending on the application. Additionally, the fact that the SVM models showed significant variation in performance in the Train, Validation and Datasets shows that these models should be implemented carefully and fully vetted before use.

For future development more observations could be included, and the data could be combined with more characteristics such as the type of phone being used, state, age, or other additional data. Additionally, other Kernels to transform the data could be attempted, and the margin of the models could be tweaked.

**References**

Han, Kamber, and Pei (2011). Data Mining: Concepts and Techniques, Third Edition Retrieved September 14, 2018 from <http://hanj.cs.illinois.edu/cs412/bk3/01.pdf>

Munnangi, H., & Chakraborty, G. (April, 2015). Predicting readmission of diabetic patients using the high performance Support Vector Machine algorithm of SAS Enterprise Miner. Retrieved from http://support.sas.com/resources/papers/proceedings15/3254-2015.pdf

Using Customer Behavior Data to Improve Customer Retention. (2015, April 11). Retrieved from <https://www.ibm.com/communities/analytics/watson-analytics-blog/predictive-insights-in-the-telco-customer-churn-data-set/>

SAS® Enterprise Miner™. (2018, November 19). Retrieved from https://support.sas.com/documentation/onlinedoc/miner/

**Appendix – SVM Analysis of Telco Data set**

Figure 1. SAS Enterprise Miner Diagram of SVM Telco Analysis

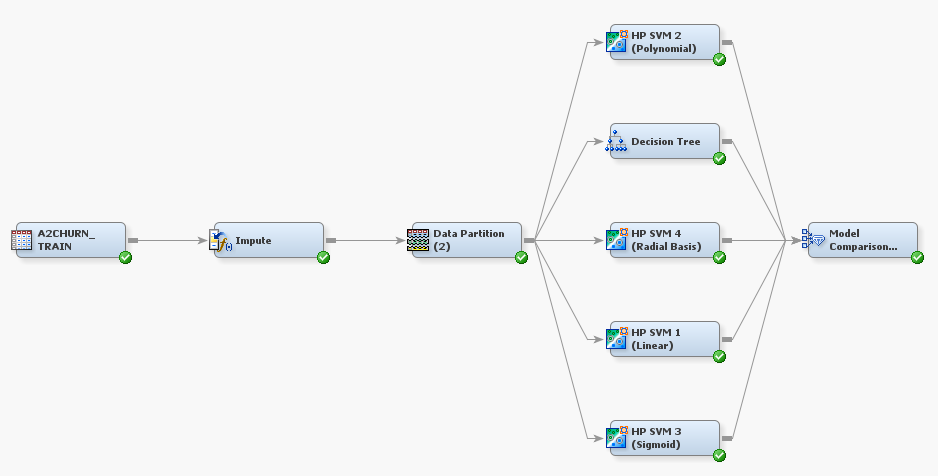
****

Table 2. Data Exploration of Telco Dataset

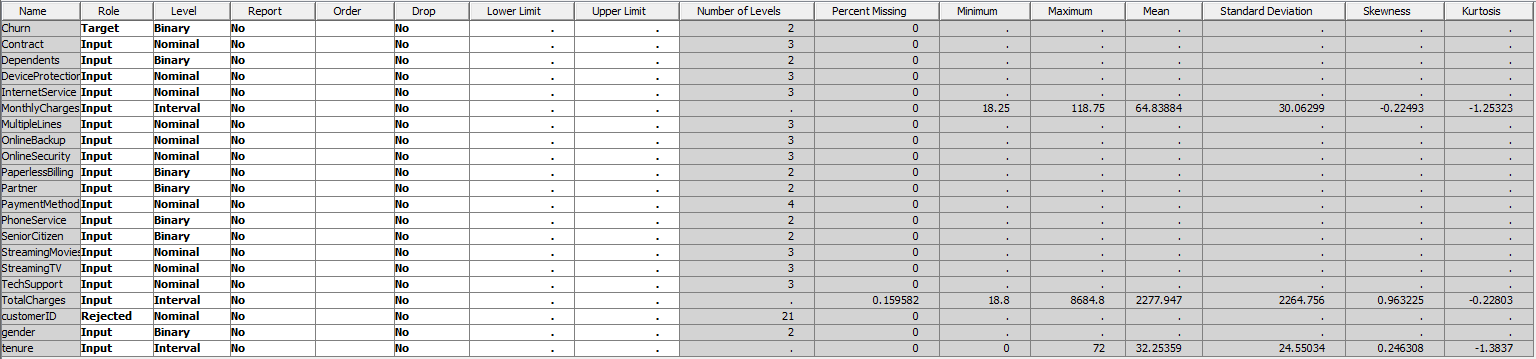
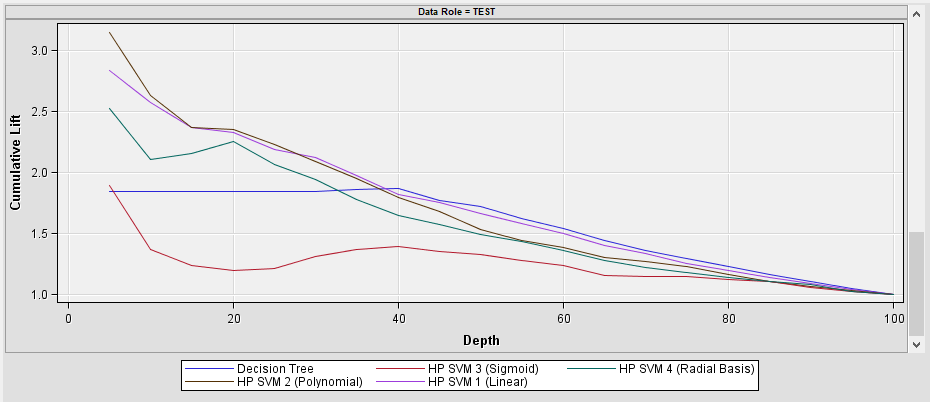
****

Figure 4. Cumulative Lift comparison

****