Churn Model Literature Review

1. An Empirical Study on Customer Churn Behaviours Prediction Using Arabic Twitter Mining Approach

This churn model was created to help solve the issue of customer churn in the voice telecommunication service industry. This team identified an issue with most churn models where the data that is most readily available is historical customer satisfaction data which does not always match or help solve current issues. So they turned to Twitter data feed to collect real-time customer satisfaction data. This presents a new challenge of text mining and turning that into useable data to feed into a model. Sentiment analysis was used to categorize tweets based on customer mood as Positive or Negative. The customer mood data combined with the structured customer data from the telecommunications company helped to create a predictive model for customer churn. The advantage of this method is that is in the collection of real time data using Twitter data and this data comes directly from the customer. The disadvantage is most people tend to only post their opinions when they are dissatisfied and satisfied customers tend to be more silent. Do you interpret the lack of feedback from Twitter users as positive by default or is there another way to measure customers who are satisfied?

2. Literature Review of Data Mining Techniques in Customer Churn Prediction for Telecommunications Industry

This churn analysis uses telecommunication data and the fuzzy algorithm for churn prediction. The fuzzy algorithm was used to clean "noisy data". This is a long paper that looks like it consists of multiple algorithms analyzed by multiple people. Each attempt had a different accuracy score but they were all in the high 90's. The conclusion does mention the use of social factors as a variable in a future model which could impact the accuracy rating. Some of the algorithms used: fuzzy algorithm, Multilayer Perceptron (MLP), "Best Hybrid Methodology", and others.

3. Customer Churn: A Study of Factors Affecting Customer Churn using Machine Learning

This churn analysis used data from IBM's “Using Customer Behavior Data to Improve Customer Retention”. Since churn analysis is limited by historical data, this researcher did their best to use the most recent data they could; data for the past month. Python was used with the package FeatureTools to create new features for analysis which resulted in 724 features from the original 19 features. The xGBoost model had the best performance in identifying churn vs not churn and the top features for this classification. LIME prediction method (Local Interpretable Model-agnostic Explanation) was also used because it makes the model easier to interpret by non-experts. Three LIME prediction models were created to display churning, non-churning, and moderate customers in a way that is easier to digest than typical model visualizations.