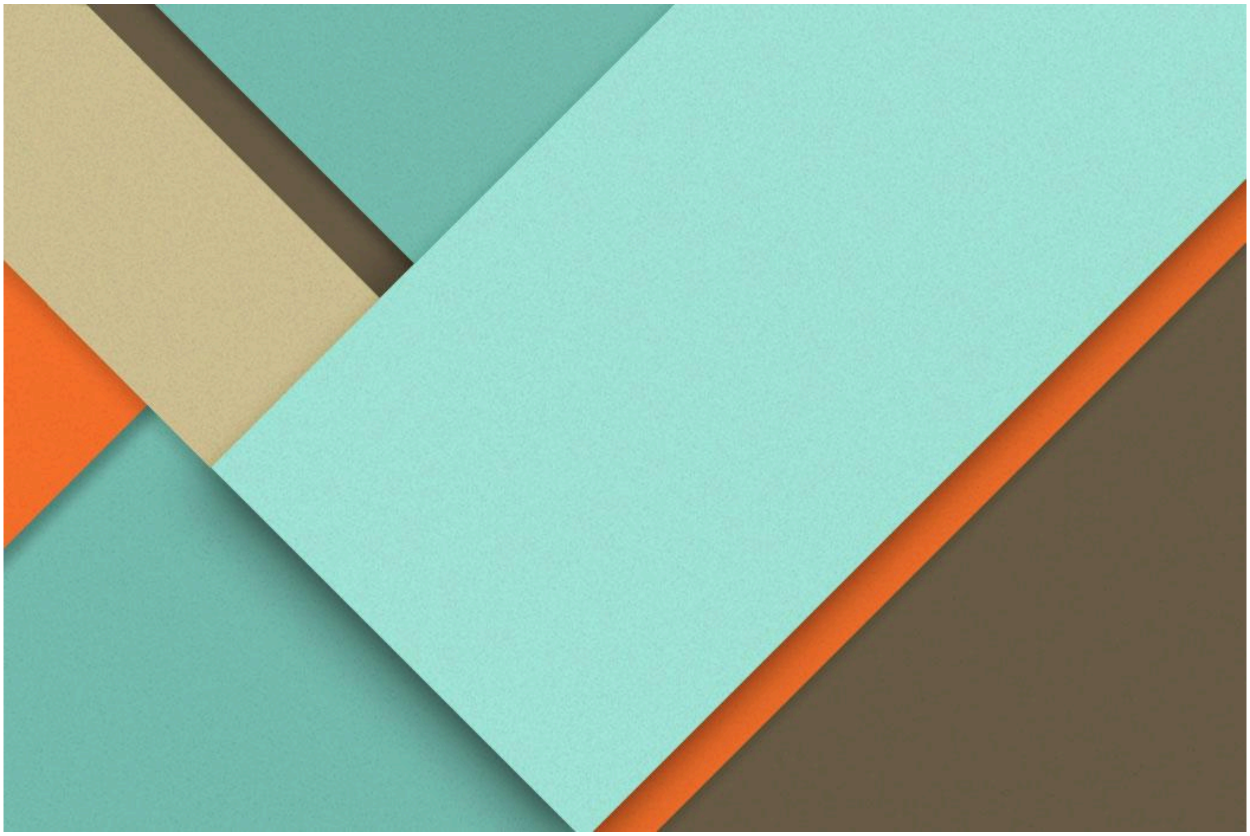

Customer Churn

01.06.2024

Mason Myre

Overview

Over the past years, the number of customers willing to pay for landline phones has decreased significantly. We believe that customers who are considering cancelling their landline phone bill will have a significant negative impact on our earnings this year. We believe that out of all the customers that plan to leave, we would benefit from simply offering them a discounted rate. This would incentivize them to stay. However, we do not want to offer that rate to all of our customers, only the ones that plan to leave. This problem seems like a great fit for the use of classification machine learning. It is important to note that if we offer a discount to a customer that planned on leaving, and still leaves despite the discount, we did not lose or gain anything by offering them a discount.

Data Set

We will be using the customer churn data set provided by coursera, as this allows me to be confident that the dataset is filled with quality data points, as opposed to data sets with potentially missing values or garbage data. The attributes are tenure, age, address, income, education, employment, equipment, callcard, wireless, and our target: churn.

I wanted to focus on getting the models to be correct as much as possible, so I went with a 75% training data, 25% testing data split as I felt that would produce the best results.

Training The Models

I decided to train 4 different models, to give me variation. Each model has their own strengths and weaknesses, so I figured training more would give me the ability to explore more of the data set and give me more advantages when the time came to pick one.

I went with Decision Tree first, as I wanted to be able to explain at least some of my findings to others, even if it didn't fit other models exactly.

I also used the random forest, bagging, and support vector models. Since these three are fairly different, it allowed me to cover more areas of what works and doesn't.

Training Results

After training the dataset on each model, I was surprised by my results. Bagging and SVC ended up tied for first at 76% test accuracy, with the Decision Tree Classifier actually beating out the Random Forest with 72% and 70% testing accuracy, respectively.

Training results for 75/25 training/test split

	SVC	Decision Tree	Random Forest	Bagging
Training Accuracy	0.78	0.90	1.00	0.97
Testing Accuracy	0.76	0.72	0.70	0.66

It is worth noting that the random forest was using Gini Index as the criteria, whereas the decision tree was using entropy, however I was still surprised to see the decision tree model outperform the random forest model. I believe my original idea of using a 75/25 training/test split was suboptimal and led to overfitting of the models. So I retried the same four models with an updated 70/30 training/test split.

Training results for the 70/30 training/test split

	SVC	Decision Tree	Bagging	Random Forest
Training Accuracy	0.79	0.86	0.96	1.00
Testing Accuracy	0.77	0.73	0.70	0.68

The results for the 70/30 split were far more surprising than the results of the 75/25. The Decision Tree model as well as The Support Vector model both experienced a small(almost negligible) improvement, whereas the Bagging model had a bit more noticeable improvement. The Random Forest model experienced a minor decrease in accuracy.

Takeaways

Overall, I would recommend using the Support Vector model for business decisions, with the decision tree as a backup to aid in the explanation of conclusions. It is not a perfect match, but can certainly help with explaining certain patterns that the SVC noticed.

From here, I would recommend two potential approaches:

1. Offer small discounts to all of the customers that the SVC model predicts to churn. We do not want to offer massive discounts, as we cannot predict the customer churn well enough and we would end up offering large discounts to some customers who were not previously considering leaving. This could also have the side effect of building goodwill with certain customers who were previously not planning on leaving but still got a discount regardless. But the value of that goodwill is something that will need to be decided on.
2. Instead of training models that only output the most likely classification, train a model to predict the probability of any given customer to churn. Then, run simulations to find the optimal probability to offer a discount at. For example: only offer a discount if the model is at least 80% sure that a client plans to no longer use our services. We could also consider offering larger discounts to clients that are more likely to churn.

Next Steps

Regardless of the approach taken, there are actions that should be taken going forward as a company to help aid our use of Artifical Intelligence and Machine Learning. These technologies are not going anywhere and if we can leverage them faster than our competitors, we will have a massive advantage in the market.

Some of these actions include:

- Getting more data to help train our models
- Sending out questionnaires to our clients
- Continuous training of models, especially as time goes on to help prevent data drift

These actions and many more will allow us to utilize Machine Learning technologies as a company, which can help predict the actions others will take, market flows, and so much more. Effective utilization can help maximize profits, and ensure that we do not fall behind in the market.