To Churn or Not to Churn?

Author: E. Thompson-Becker

Date: August 26th, 2022

Table of Contents

Purpose

Objective

- Telephone company wants to predict if a customer will churn or not
- The company wants to know what is the best classification algorithm to use to fulfill their needs
  - Naïve Bayes Classification
  - Decision Tree Classification

The Data

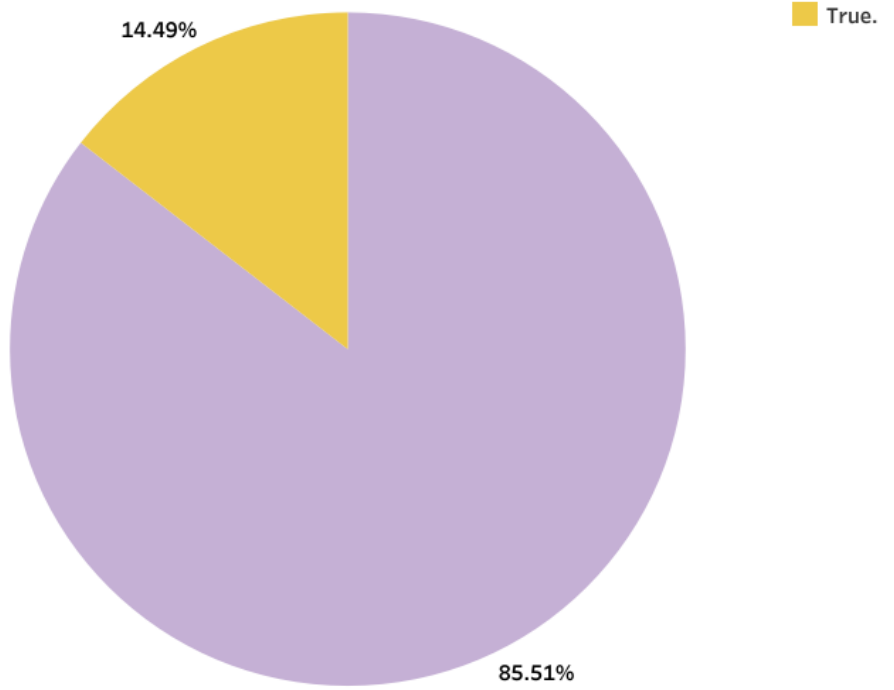
The Data

Attribute	Data Type	Description
State	Chr	Customer's State
Account Length	Numeric	Duration of activity for a customer account
Area Code	Numeric	Area Code for customer
Phone	Chr	Customer's Phone Number
Intl Plan	Chr (yes/no)	Indicator if customer has an international plan
Vmail Plan	Chr (yes/no)	Indicator of a voice mail plan
Vmail Message	Numeric	No. of voicemail messages
Day Mins	Numeric	No. of minutes the service used during the day
Day Calls	Numeric	No. of calls during the day time
Day Charge	Numeric	Charges for using the service during the day time
Eve Mins	Numeric	No. of minutes of service used during the evening

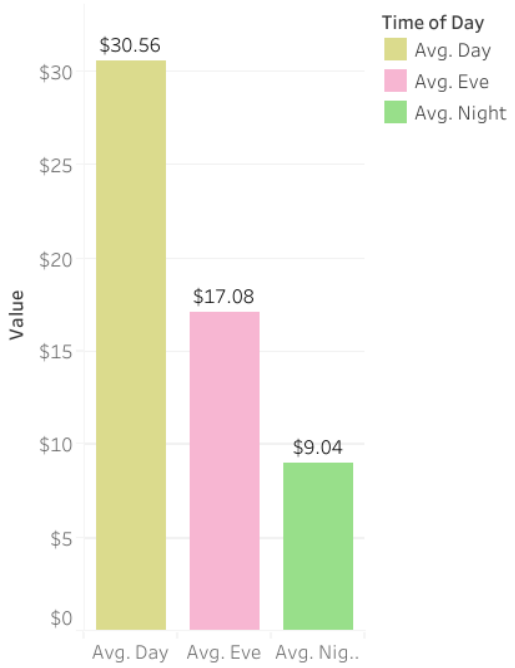
Attribute	Data Type	Description
Eve Calls	Numeric	No. of calls during the evening time
Eve Charge	Numeric	Charges for using the service during the evening
Night Mins	Numeric	No. of minutes of service used during the night time
Night Calls	Numeric	No. of calls made during the night time
Night Charge	Numeric	Charges for using the service during the night time
Intl Mins	Numeric	No. of minutes used to make international calls
Intl Calls	Numeric	No. of international calls
Intl Charge	Numeric	Charges for international calls
CustServ Calls	Numeric	No. of calls to customer support service
Churn?	Chr (True/False)	Class attribute with binary values

The Data

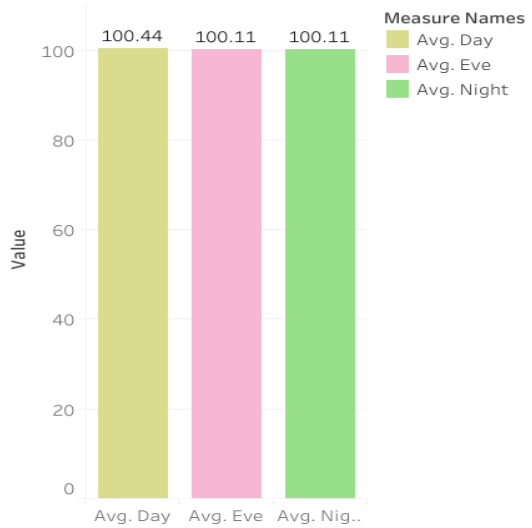
Percentage of Churn



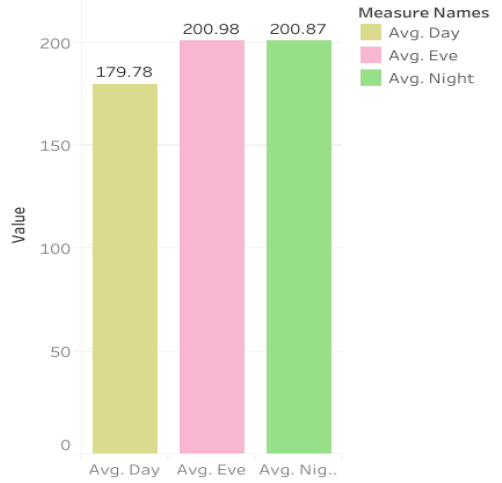
Average Charge by Time of Day



Average Number of Calls by Time of Day



Average Minutes by Time of Day



Classification

Decision Tree Classification

Results

Decision tree classification with all attributes:
93% correctly classified
7% incorrectly classified

Baseline	False	True
False	830	26
True	44	100

Baseline	False	True
Precision	94.97	79.37
Recall	96.96	69.44

Decision tree classification with selected attributes:
94.7 % correctly classified
5.3% incorrectly classified

Selected	False	True
False	847	9
True	44	100

Selected	False	True
Precision	95.06	91.74
Recall	98.95	69.44

Naïve Bayes Classification

Results

Naïve Bayes classification with all attributes:
87.7 % correctly identified
12.3% incorrectly identified

Baseline	False	True
False	803	53
True	70	74

Baseline	False	True
Precision	91.98	58.27
Recall	93.81	51.39

Naïve Bayes classification with selected attributes:
87.2% correctly identified
12.8% incorrectly identified

Selected	False	True
False	831	25
True	103	41

Selected	False	True
Precision	95.06	91.74
Recall	98.95	69.44

Comparison

Evaluation Metrics

$$F\_Measure = \frac{(2 * Precision * Recall)}{(Precision + Recall)}$$

Model	F-Measure
Decision-Tree-Baseline	0.741
Decision-Tree-Selected	0.791
Naïve-Bayes-Baseline	0.546
Naïve-Bayes-Selected	0.390

Conclusion

Which method is better?

Decision Tree Classification

Combining the f-measure value and the accuracy measurement, the best model was the selected features decision tree classification model.

With an accuracy of 94.7% correctly identified cases and an f-measure of 0.791.