

Case Study - Customer Churn

John D. Kelleher and Brian Mac Namee and Aoife D'Arcy

- 1 **Business Understanding**
- 2 **Data Understanding**
- 3 **Data Preparation**
- 4 **Modelling**
- 5 **Evaluation**
- 6 **Deployment**

- Acme Telephonica (AT) is a mobile phone operator that has customers across every state of the U.S.A.
- AT struggles with customer **churn prediction**—customers leaving AT for other mobile phone operators.
- In 2010 AT hired Ross, a predictive data analytics specialist, to take a new approach to reducing customer churn.
- This case study describes the work carried out by Ross when he took AT through the CRISP-DM process to develop a predictive data analytics solution to this business problem.

Business Understanding

- AT did not approach Ross with a well-specified predictive analytics solution. Instead, the company approached him with a business problem—reducing customer churn.
- Ross's first goal was to convert this business problem into a concrete analytics solution.
- Before attempting this conversion, Ross had to fully understand the business objectives of AT.

Data Understanding

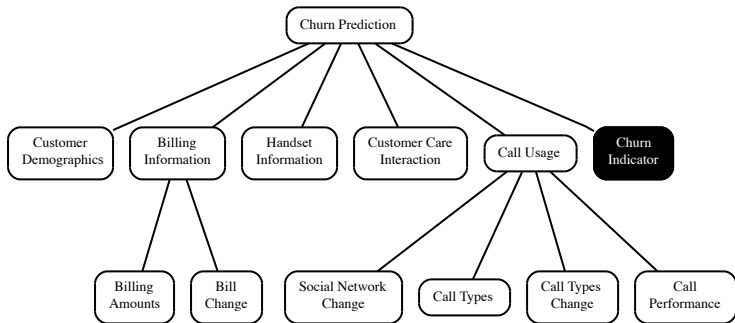


Figure: The set of domain concepts for the Acme Telephonica customer churn prediction problem.

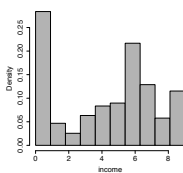
Feature	Description
BILLAMOUNTCHANGE PCT	The percent by which the customer's bill has changed from last month to this month
CALLMINUTESCHANGE PCT	The percent by which the call minutes used by the customer has changed from last month to this month
AVGBILL	The average monthly bill amount
AVGRECURRINGCHARGE	The average monthly recurring charge paid by the customer
AVGDROPPEDCALLS	The average number of customer calls dropped each month
PEAKRATIOCHANGE PCT	The percent by which the customer's peak calls to off-peak calls ratio has changed from last month to this month
AVGRECEIVEDMINS	The average number of calls received each month by the customer
AVGMINS	The average number of call minutes used by the customer each month
AVGOVERBUNDLEMINS	The average number of out-of-bundle minutes used by the customer each month
AVGROAMCALLS	The average number of roaming calls made by the customer each month
PEAKOFFPEAKRATIO	The ratio between peak and off peak calls made by the customer this month
NEWFREQUENTNUMBERS	How many new numbers the customer is frequently calling this month?

Feature	Description
CUSTOMERCARECALLS	The number of customer care calls made by the customer last month
NUMRETENTIONCALLS	The number of times the customer has been called by the retention team
NUMRETENTIONOFFERS	The number of retention offers the customer has accepted
AGE	The customer's age
CREDITRATING	The customer's credit rating
INCOME	The customer's income level
LIFETIME	The number of months the customer has been with AT
OCCUPATION	The customer's occupation
REGIONTYPE	The type of region the customer lives in
HANDSETPRICE	The price of the customer's current handset
HANDSETAGE	The age of the customer's current handset
NUMHANDSETS	The number of handsets the customer has had in the past 3 years
SMARTPHONE	Is the customer's current handset a smart phone?
CHURN	The target feature

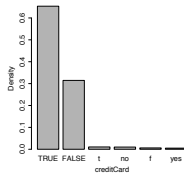
Data Preparation

Feature	Count	% Miss.	Card.	Min.	1 st Qrt.	Mean	Median	3 rd Qrt.	Max.	Std. Dev.
AGE	10,000	11.47	40	0.00	0.00	30.32	34.00	48.00	98.00	22.16
INCOME	10,000	0.00	10	0.00	0.00	4.30	5.00	7.00	9.00	3.14
NUMHANDSETS	10,000	0.00	19	1.00	1.00	1.81	1.00	2.00	21.00	1.35
HANDSETAGE	10,000	0.00	1,923	52.00	590.00	905.52	887.50	1,198.00	2,679.00	453.75
HANDSETPRICE	10,000	0.00	16	0.00	0.00	35.73	0.00	59.99	499.99	57.07
AVGBILL	10,000	0.00	5,588	0.00	33.33	58.93	49.21	71.76	584.23	43.89
AVGMIN	10,000	0.00	4,461	0.00	150.63	521.17	359.63	709.19	6,336.25	540.44
AVGRECURRINGCHARGE	10,000	0.00	1,380	0.00	30.00	46.24	44.99	59.99	337.98	23.97
AVGOVERBUNDLEMIN	10,000	0.00	2,808	0.00	0.00	40.65	0.00	37.73	513.84	81.12
AVGROAMCALLS	10,000	0.00	850	0.00	0.00	1.19	0.00	0.26	177.99	6.05
CALLMINUTESCHANGEPC	10,000	0.00	10,000	-16.422	-1.49	0.76	0.50	2.74	19.28	3.86
BILLAMOUNTCHANGEPC	10,000	0.00	10,000	-31.67	-2.63	2.96	1.96	7.56	42.89	8.51
AVGRECEIVEDMIN	10,000	0.00	7,103	0.00	7.69	115.27	52.54	154.38	2,006.29	169.98
AVGOUTCALLS	10,000	0.00	524	0.00	3.00	25.29	13.33	33.33	610.33	35.66
AVGINCALLS	10,000	0.00	310	0.00	0.00	8.37	2.00	9.00	304.00	17.68
PEAKOFFPEAKRATIO	10,000	0.00	8,307	0.00	0.78	2.22	1.40	2.50	160.00	3.88
PEAKOFFPEAKRATIOCHANGEPC	10,000	0.00	10,000	-41.32	-6.79	-0.05	0.01	6.50	37.78	9.97
AVGDROPPEDCALLS	10,000	0.00	1,479	0.00	0.00	0.50	0.00	0.00	9.89	1.41
LIFETIME	10,000	0.00	56	6.00	11.00	18.84	17.00	24.00	61.00	9.61
LASTMONTHCUSTOMERCARECALLS	10,000	0.00	109	0.00	0.00	1.74	0.00	1.33	365.67	5.76
NUMRETENTIONCALLS	10,000	0.00	5	0.00	0.00	0.05	0.00	0.00	4.00	0.23
NUMRETENTIONOFFERSACCEPTED	10,000	0.00	5	0.00	0.00	0.02	0.00	0.00	4.00	0.155
NEWFREQUENTNUMBERS	10,000	0.00	4	0.00	0.00	0.20	0.00	0.00	3.00	0.64

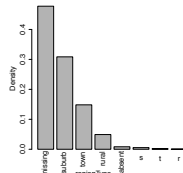
								2 nd	2 nd
Feature	Count	% Miss.	Card.	Mode	Mode Freq.	Mode %	2 nd Mode	Mode Freq.	Mode %
OCCUPATION	10,000	74.00	8	professional	1,705	65.58	crafts	274	10.54
REGIONTYPE	10,000	47.80	8	suburb	3,085	59.05	town	1,483	28.39
MARRIAGESTATUS	10,000	0.00	3	unknown	3,920	39.20	yes	3,594	35.94
CHILDREN	10,000	0.00	2	FALSE	7,559	75.59	TRUE	2,441	24.41
SMARTPHONE	10,000	0.00	2	TRUE	9,015	90.15	FALSE	985	9.85
CREDITRATING	10,000	0.00	7	B	3,785	37.85	C	1,713	17.13
HOMEOWNER	10,000	0.00	2	FALSE	6,577	65.77	TRUE	3,423	34.23
CREDITCARD	10,000	0.00	6	TRUE	6,537	65.37	FALSE	3,146	31.46
CHURN	10,000	0.00	2	FALSE	5,000	50.00	TRUE	5,000	50.00



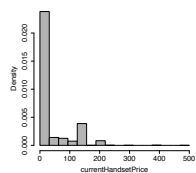
(a) INCOME



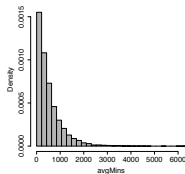
(b) CREDITCARD



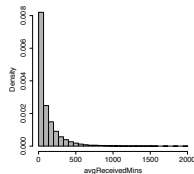
(c) REGIONTYPE



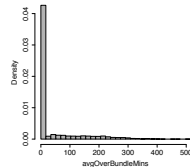
(d) HANDSET-PRICE



(e) AVGMINS

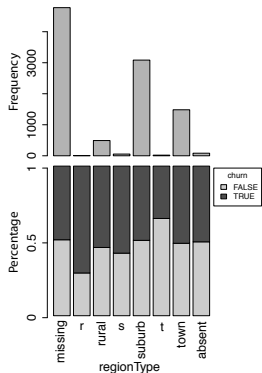


(f) AVGRECEIVED-MINS

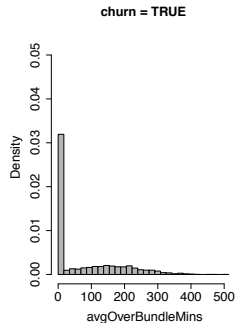
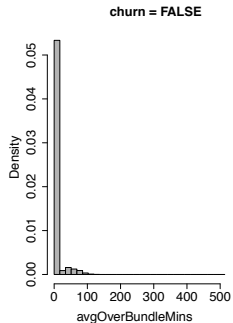


(g) AVGOVER-BUNDLEMINS

Figure: (a) - (c) Histograms for the features from the AT ABT with irregular cardinality. (d) - (g) Histograms for the features from the AT ABT that are potentially suffering from outliers.



(a) REGIONTYPE



(b) AVGOVERBUNDLEMINS

Figure: (a) a stacked bar plot for the REGIONTYPE FEATURE. (b) histograms for the AVGOVERBUNDLEMINS feature by target feature value.

Modelling

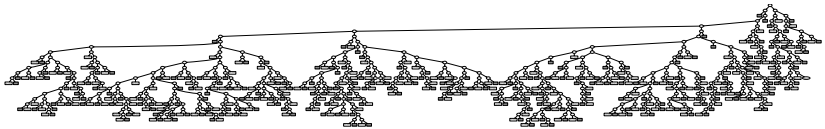


Figure: An unpruned decision tree built for the AT churn prediction problem (shown only to indicate its size and complexity). The excessive complexity and depth of the tree are evidence that over-fitting has probably occurred.

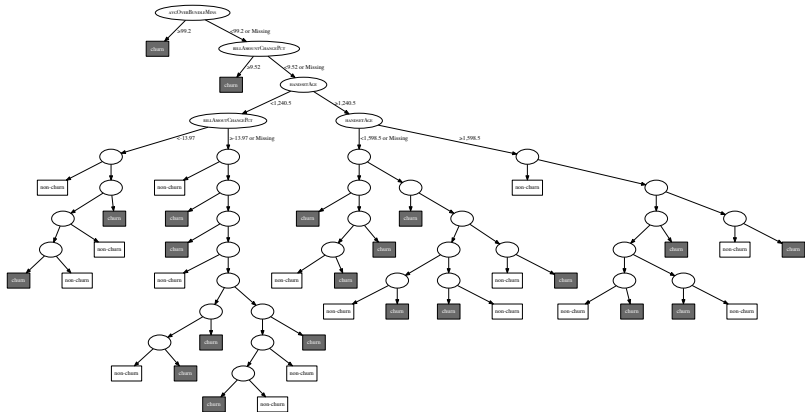


Figure: A pruned decision tree built for the AT churn prediction problem. Grey leaf nodes indicate a churn prediction while clear leaf nodes indicate a non-churn prediction. For space reasons we only show the features tested at the top level nodes.

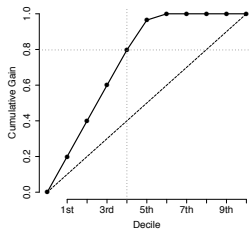
Table: The confusion matrix from the test of the AT churn prediction stratified hold-out test set using the pruned decision tree in Figure 5 [17].

		Prediction		Recall
		'churn'	'non-churn'	
Target	'churn'	1,058	442	70.53
	'non-churn'	152	1,348	89.86

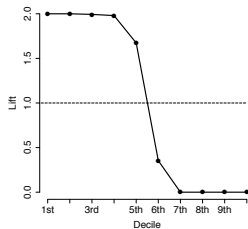
Evaluation

Table: The confusion matrix from the test of the AT churn prediction non-stratified hold-out test set.

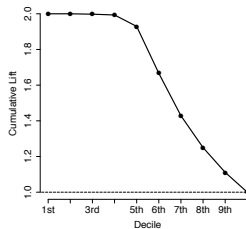
		Prediction		Recall
		'churn'	'non-churn'	
Target	'churn'	1,115	458	70.88
	'non-churn'	1,439	12,878	89.95



(a) Cumulative gain



(b) Lift



(c) Cumulative lift

Figure: (a) cumulative gain, (b) lift and (c) cumulative lift charts for the predictions made on the large test data sample.

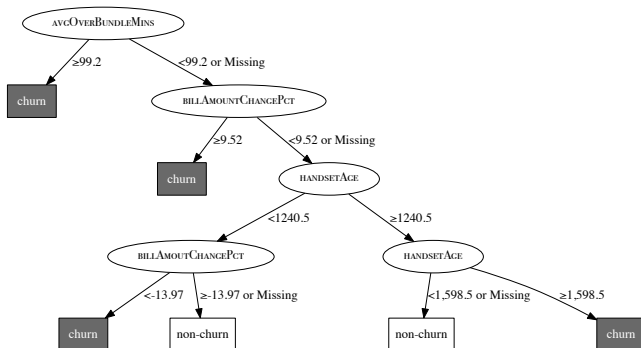


Figure: A pruned and stunted decision tree built for the Acme Telephonica churn prediction problem

Deployment

- Because AT was already using a process in which its retention team generated call lists based on collected data, deployment of the new decision tree model was reasonably straight-forward.
- Ross worked with the AT IT department to develop deployment-ready **extract-transform-load** (ETL) routines to generate queries for the model.
- The last step in deployment was to put in place an **on-going model validation** plan to raise an alarm if evidence arose indicating that the deployed model had gone **stale**.

- 1 **Business Understanding**
- 2 **Data Understanding**
- 3 **Data Preparation**
- 4 **Modelling**
- 5 **Evaluation**
- 6 **Deployment**