MegaTelCo: Predicting Customer Churn

You just landed a great analytical job with MegaTelCo, one of the largest telecommunication firms in the US.

They are having a major problem with customer retention in their wireless business:

- In the mid-Atlantic region, 20% of cell phone customers leave when their contracts expire. Communications companies are now engaged in battles to attract each other’s customers while retaining their own.

- Marketing has already designed a special retention offer.

- Your task is to devise a precise, step-by-step plan for how the data science team should use MegaTelCo’s vast data resources to solve the problem.
MegaTelCo: Predicting Customer Churn

• What data you might use?

• How would they be used?

• How should MegaTelCo choose a set of customers to receive their offer in order to best reduce churn for a particular incentive budget?
Terminology

• Model:
  • A simplified representation of reality created to serve a purpose

• Predictive Model:
  • A formula for estimating the unknown value of interest: the target
    • The formula can be mathematical, logical statement (e.g., rule), etc.

• Prediction:
  • Estimate an unknown value (i.e. the target)

• Instance / example:
  • Represents a fact or a data point
  • Described by a set of attributes (fields, columns, variables, or features)
Terminology

• Model induction:
  • The creation of **models** from data

• Training data:
  • The input data for the induction algorithm
## Terminology

![Table and diagram showing attributes and a target attribute.](image-url)

<table>
<thead>
<tr>
<th>Name</th>
<th>Balance</th>
<th>Age</th>
<th>Employed</th>
<th>Write-off</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mike</td>
<td>$200,000</td>
<td>42</td>
<td>no</td>
<td>yes</td>
</tr>
<tr>
<td>Mary</td>
<td>$35,000</td>
<td>33</td>
<td>yes</td>
<td>no</td>
</tr>
<tr>
<td>Claudio</td>
<td>$115,000</td>
<td>40</td>
<td>no</td>
<td>no</td>
</tr>
<tr>
<td>Robert</td>
<td>$29,000</td>
<td>23</td>
<td>yes</td>
<td>yes</td>
</tr>
<tr>
<td>Dora</td>
<td>$72,000</td>
<td>31</td>
<td>no</td>
<td>no</td>
</tr>
</tbody>
</table>

This is one row (example). Feature vector is: `<Claudio,115000,40,no>`
Class label (value of Target attribute) is **no**
What is a model?

A simplified* representation of reality created for a specific purpose

*based on some assumptions

• Examples: map, prototype, Black-Scholes model, etc.

• Data Mining example:
  “formula” for predicting probability of customer attrition at contract expiration
  → “classification model” or “class-probability estimation model”
Feature Types

• Numeric: anything that has some order
  ▪ Numbers (that mean numbers)
  ▪ Dates (that look like numbers …)
  ▪ Dimension of 1

• Categorical: stuff that does not have an order
  ▪ Binary
  ▪ Text
  ▪ Dimension = number of possible values (-1)

• Food for thought: Names, Ratings, SIC
Dimensionality of the data?

**Attributes / Features**

<table>
<thead>
<tr>
<th>Name</th>
<th>Balance</th>
<th>Age</th>
<th>Default</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mike</td>
<td>$123,000</td>
<td>30</td>
<td>Yes</td>
</tr>
<tr>
<td>Mary</td>
<td>$51,100</td>
<td>40</td>
<td>Yes</td>
</tr>
<tr>
<td>Bill</td>
<td>$68,000</td>
<td>55</td>
<td>No</td>
</tr>
<tr>
<td>Jim</td>
<td>$74,000</td>
<td>46</td>
<td>No</td>
</tr>
<tr>
<td>Mark</td>
<td>$23,000</td>
<td>47</td>
<td>Yes</td>
</tr>
<tr>
<td>Anne</td>
<td>$100,000</td>
<td>49</td>
<td>No</td>
</tr>
</tbody>
</table>

- Dimensionality of a dataset is the sum of the dimensions of the features
- The sum of the number of numeric features and ~ the number of values of categorical features
Common Data Mining Tasks

• Classification and class probability estimation
  • How likely is this consumer to respond to our campaign?

• Regression
  • How much will she use the service?

• Similarity Matching
  • Can we find consumers similar to my best customers?

• Clustering
  • Do my customers form natural groups?
Common Data Mining Tasks

- **Co-occurrence Grouping**
  - Also known as frequent itemset mining, association rule discovery, and market-basket analysis
  - What items are commonly purchased together?

- **Profiling (behavior description)**
  - What does “normal behavior” look like? (for example, as baseline to detect fraud)

- **Data Reduction**
  - Which latent dimensions describe the consumer taste preferences?
Common Data Mining Tasks

• Link Prediction
  • Since John and Jane share 2 friends, should John become Jane’s friend?

• Causal Modeling
  • Why are my customers leaving?
Supervised versus Unsupervised Methods

- “Do our customers naturally fall into different groups?”
  - No guarantee that the results are meaningful or will be useful for any particular purpose

- “Can we find groups of customers who have particularly high likelihoods of canceling their service soon after contracts expire?”
  - A specific purpose
  - Much more useful results (usually)
  - Different techniques
  - Requires data on the target
    - The individual’s label
## Common Data Mining Tasks

<table>
<thead>
<tr>
<th>Task</th>
<th>Supervised Methods</th>
<th>Unsupervised Methods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Regression</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Causal Modeling</td>
<td>✓</td>
<td></td>
</tr>
<tr>
<td>Similarity Matching</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Link Prediction</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Data Reduction</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Clustering</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Co-occurrence Grouping</td>
<td></td>
<td>✓</td>
</tr>
<tr>
<td>Profiling</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>
Supervised Data Mining & Predictive Modeling

• Is there a specific, quantifiable target that we are interested in or trying to predict?
  • Think about the decision

• Do we have data on this target?
  • Do we have enough data on this target?
    • Need a min ~500 of each type of classification

• Do we have relevant data prior to decision?
  • Think timing of decision and action

• The result of supervised data mining is a model that predicts some quantity

• A model can either be used to predict or to understand
Subclasses of Supervised Data Mining

- Classification
  - Categorical target
    - Often binary
  - Includes “class probability estimation”

- Regression
  - Numeric target
Subclasses of Supervised Data Mining

- “Will this customer purchase service $S_1$ if given incentive $I_1$?”
  - Classification problem
    - Binary target (the customer either purchases or does not)

- “Which service package ($S_1$, $S_2$, or none) will a customer likely purchase if given incentive $I_1$?”
  - Classification problem
    - Three-valued target

- “How much will this customer use the service?”
  - Regression problem
    - Numeric target
      - Target variable: amount of usage per customer
**Data Mining versus Use of the Model**

"Supervised" modeling:

- **Data**: Database containing historical data.
- **Data Mining**: Process of extracting patterns from data.
- **Model**: Predictive model generated from data mining.
- **Prediction**: Output of the model.
- **New data item**: An item with some value unknown (e.g., will she leave?).

"Training" data have all values specified.

New data item has some value unknown (e.g., will she leave?)
Classical Pitfalls in DM setup

• The training data is NOT consistent with the use

• Bad sample

• Bad features
Sample: “Looking under the streetlight”

- Target Proxy
  - I do not see if a person after seeing an ad bought the book, so let's model clicks …

- Sample Proxy
  - I want to run a campaign in Spain but only have data on US customers
Sample: “Survivorship issues”

• Lending club wants to have a model to take over the screening process that selects applications and deny those that are likely to default

• Data of past loans and the outcomes are provided

• Bad:
  ▪ Use the data they currently have to predict default
**Sample: Different Sources**

- Things go really bad if the positive and negative are treated differently

- Looking for drivers of diabetes: how do you assemble the training data?

  - Bad:
    - Go to a specialized hospital and get records from people treated for diabetes
    - Go somewhere else to get records for healthy people
Summary on bad habits

• You are missing all the applications that were turned down already
• The sick people came from a very artificial subset
• Your target is NOT really your target
• No way of telling how the model will perform
  ▪ No way of testing either

The training sample should be as similar as possible to the USE data
Digression on features: It is all about the timing in use!

$T_d$: Decision has to be made

$T_i$: Target is known

Features  Leakage
Thanks!
Questions?