Understanding proposal win rates

Martijn Schroder

1/06/2019

Contents

Executive summary

In management consulting competitiveness is high. Our firm competes actively with the 'Big 4' consultancies. Better understanding of the reasons why we win and lose proposal will give us an advantage.

Our business makes decisions on our proposal management practice with the view to increase win rates. Until recently those decisions were made based on experience and perception of what works and what doesn't.

In this assignment my aim is to apply manchine learning approaches to gain insights into what the data tells us about the relevant features that are good predictors of win and lose rates.

Given the low number of transactions and limited cleanliness of the data, the analysis of features that should underpin decisions around proposals is relatively ambiguous. This is the main challenge to work with.

Cleaning data

The data set is a raw export from the system we use to manage opportunities. A csv export was obtained with the following structure:

```r
names(proposals) # obtain the column names
```

```
##  [1] "Opportunity Name"
##  [2] "Account Name"
##  [3] "Stage"
##  [4] "Amount Currency"
##  [5] "Amount"
##  [6] "Created Date"
##  [7] "Close Date"
##  [8] "Primary Practice"
##  [9] "Business Offer"
## [10] "Sector"
```

```
## [11] "Segment"
## [12] "Proposal director"
## [13] "Proposal manager"
## [14] "Source"
## [15] "Competitive or sole sourced (compulsory)"
```

```r
tibble(proposals) # show structure of data
```

```
## # A tibble: 3,562 x 1
##    proposals$`Oppo~ $`Account Name` $Stage $`Amount Curren~ $Amount
##    <chr>            <chr>           <chr>  <chr>              <dbl>
##  1 USY 1803 Sydney~ University of ~ Clien~ AUD              300000
##  2 UOW 1803 Academ~ University of ~ Opp s~ AUD               81735
##  3 NAU 1803 - HSF ~ Nous Australia  Opp s~ GBP                   1
##  4 DHV 1608 Improv~ Dept of Health~ Opp s~ AUD               99134.
##  5 NUK 1701 - CCS ~ Nous UK         Opp s~ AUD                   0
##  6 DHG 1803 NW Abo~ Dept of Commun~ Clien~ AUD               15000
##  7 UNS 1803 Refine~ University of ~ Opp s~ AUD              220000
##  8 UNS 1803 Market~ University of ~ Opp u~ AUD              150000
##  9 LTU 1803 Accomm~ La Trobe Unive~ Opp u~ AUD              125000
## 10 DTF 1802 Major ~ Dept of Treasu~ Opp u~ AUD              300000
## # ... with 3,552 more rows, and 10 more variables: $`Created Date` <chr>,
## #   $`Close Date` <chr>, $`Primary Practice` <chr>, $`Business
## #   Offer` <chr>, $Sector <chr>, $Segment <chr>, $`Proposal
## #   director` <chr>, $`Proposal manager` <chr>, $Source <chr>,
## #   $`Competitive or sole sourced (compulsory)` <chr>
```

The data needs cleaning up. The following changes are made:

- Rename the columns with names more suitable for analysis
- Factorise features to enable certain type analyses
- Convert dates and proposal amounts from strings to dates and integers
- To facilitate Principle Component Analysis (PCA) columns reflecting people directing and managing the opportunities will be converted using one hot encoding. This will be done after data exploration and at the appropriate stage of analysis

```r
# Rename columns to increase readibility and facilitate analysis
proposals <- proposals %>%
  rename(
    name = `Opportunity Name`,
    account = `Account Name`,
    stage = Stage,
    currency = `Amount Currency`,
    amount = Amount,
    practice = `Primary Practice`,
    offer = `Business Offer`,
    sector = Sector,
    director = `Proposal director`,
    manager = `Proposal manager`,
    source = Source,
    competitiveness = `Competitive or sole sourced (compulsory)`,
    amount = Amount,
    segment = Segment,
    creationDate = `Created Date`,
    closeDate = `Close Date`
  )
```

```r
# Factorise culumns to facilitate certain type analyses (with exception of regression)
proposals$account <- factor(proposals$account)
proposals$stage <- factor(proposals$stage)
proposals$currency <- factor(proposals$currency)
proposals$practice <- factor(proposals$practice)
proposals$offer <- factor(proposals$offer)
proposals$sector <- factor(proposals$sector)
proposals$director <- factor(proposals$director)
proposals$manager <- factor(proposals$manager)
proposals$source <- factor(proposals$source)
proposals$competitiveness <- factor(proposals$competitiveness)
proposals$segment <- factor(proposals$segment)

# Convert amount column from char to integer
proposals$amount <- as.integer(proposals$amount)

# Convert date columns from char to date types
proposals$creationDate <- as.Date(proposals$creationDate, "%d/%m/%Y")
proposals$closeDate <- as.Date(proposals$closeDate, "%d/%m/%Y")
```

Exploration

Data integrity

The following table shows NA values in the dataset. It is clear that some work is to be done to prepare the dataset for analysis.

```r
# Show a table summary of the NA values in the dataset
colSums(is.na(proposals))
```

```
##            name         account           stage        currency
##               2               7               7               7
##          amount    creationDate       closeDate        practice
##              82               7               7             118
##           offer          sector         segment        director
##               9            1693             604              22
##         manager          source competitiveness
##              24            1328            3331
```

The following approaches to addressing the NA values are proposed.

- Drop the column "Competitive or sole sourced (compulsory)". Although it would be interesting to see the impact of competitiveness on proposals, too many data points are missing to make it useful
- Amount: investigate if mmissing amounts can be replaced with amount group means for "account", "primary practice" and "business offer"
- Proposal director / manager: create "unknown" category for relevant observations. Doing this will retain observations for analysis and will not interfere with PCA
- Outline other wrangling to be conducted

```r
# Drop "Competitive or sole sourced (compulsory)" column
proposals <- select(proposals,-competitiveness)
```

Method

Analysis

Conclusions