Predicting Sales Win or Lose

Herman Toeante

February 12, 2019

Introduction

How do you know which deals will close? You've worked your territory and leads. Demonstrated the value of your product or service to your customers' businesses and it should be a done deal. But is it? All too often, determining which customers will buy is a guessing game. You have pipeline reports, regional sales figures and wins and losses that you could analyze. Unlock the information in those sources and you'll unlock more revenue and more satisfied customers. Better understanding of sales pipeline can help any sales team organization can expect win or lose based on data. In this project, I am going to be the sales manager at an automotive supply company. I can apply similar approach to any B2B company like GE, AMAZON, GOOGLE, ADP, etc. As a manager, I'm trying to assess a sales execution issue. We have not been able to convert enough opportunities lately.

Data

Data for this project is publicly available from Wattson Analytic sample data. (https://community.watsonanalytics.com/wp-content/uploads/2015/04/WA_Fn-UseC_-Sales-Win-Loss.csv) The csv file contains 78K row of data and 19 columns. Each row is assigned a unique sales opportunity ID. The dependent variable is 'Opportunity Result' column with values of either 'Won' or 'Loss' There are several independent variables from the sample data that I can use such as: 'Supplies Group', 'Region', 'Route To Market', 'Elapsed Days In Sales Stage', 'Opportunity Result', 'Sales Stage Change Count', 'Total Days Identified Through Closing', 'Total Days Identified Through Qualified', 'Opportunity Amount USD', 'Client Size By Revenue', 'Client Size By Employee Count', 'Revenue From Client Past Two Years', and 'Competitor Type.'

Approach

Apart from doing some data ad-hoc exploration, I'll use the predictive Analytics to understand the most important elements that go into a successful deal. In the data exploration stage, I will use bar chart, scater chart, stack bar chart, and column stack bar chart. To understand the important independent variables affecting sucessfule deal, I will use linear regression, logistic regression, k-cluster regression, and random forest regression methods.

Wrangling

A minimum amount of wrangling was necessary in order to prepare the data set. The sample data mentioned above were imported and minor changes were made to the column headers.