

# GT Introduction to Analytics Modeling - Course Project

In this course project, we read and think about what analytics models and data might have been required. Specifically, we review a project for which at least three different analytics models might have been combined to create the solution.

Upon selecting an article, we describe what models might be used to create the solution, how they would be combined, what specific data might be needed to use the models, how it might be collected, and how often it might need to be refreshed and the models re-run.

## Selected Article Summary

For this assignment, we selected the article titled “**Analytics for the Engagement Life Cycle of IBM’s Highly Valued IT Service Contracts**” located at <https://www.informs.org/Impact/O.R.-Analytics-Success-Stories/Analytics-for-the-Engagement-Life-Cycle-of-IBM-s-Highly-Valued-IT-Service-Contracts>. This article describes a suite of analytics tools that were created to bring efficiencies to the process of responding to customer RFPs.

There’s a lot that could be covered in this article. For this assignment, we’ll focus on items described in the article as follows:

1. A requirement analysis tool that uses innovative NLP techniques and data mining to analyze clients’ RFP documents, extract requirements, and map them to relevant offerings.
2. A pricing tool that consists of data mining of historical deals and market data in order to calculate pricing points, and a predictive model that provides the relative win probability of each price point.

## Models, Data, and Maintenance

The 2 items above could be implemented with the following models. We also describe data and model maintenance.

### Mapping Content to Offerings

Both of our content sources (the RFP and Offerings) exist as text documents. Using NLP techniques (as mentioned in the article) we could decompose all of the documents into a Bag of Words structure. Once in this structure, we can use a similarity measure such as cosine similarity to determine which Offerings are the closest match to the RFP. Regarding model maintenance, we would need to remove obsolete offerings from the model, and add new offerings to the model.

As an alternative, past RFP documents could be manually mapped to Offerings to generate labeled data. This labeled data could then be used to train a logistic regression model that provides a probability of mapping to a particular offering. Since we have multiple labels (Offerings), this would require a multinomial approach. Regarding model maintenance, this model would be updated as RFPs are made available.

### Predicting Pricing Points

Predicting pricing points could be achieved using a regression model. As mentioned in the article, data mining of historical deals and market data would lead to the development of features. I would assume that since we are analyzing historical deals, we know the price points of those deals, and therefore the historical data is labeled. This labeled data could then be used to train, test, and validate a regression model that could be used to predict pricing points of new RFPs.

This model would likely need to be updated regularly as prices change due to technology changes, economic conditions, and competition.

### **Predicting Probability of a Win**

Logistic regression models provide a probability measure. As with price point prediction, the data mining of historical deals would lead to the development of features. And since this data is historical, the historical data would be labeled with successes and losses. This labeled data could then be used to train, test, and validate a logistic regression model that could be used to provide a probability of winning. Regarding model maintenance, this model would be updated as additional RFPs are made available.

### **Combining Models**

3 common questions that an Information Technology vendor will ask when competing for business are:

1. Can we perform the work?
2. Can we win the bid?
3. Can we make money doing it?

The models described above can help answer these questions. The first can be forecasted by the mapping between the RFP and our offerings. The second can be forecasted by or logistic regression model predicting wins. The third can be derived from our pricing point forecast.

As an additional thought, perhaps we can combine these in a decision tree. Each decision node would be an answer to one of the 3 questions. The first is likely a binary decision based on mapping to an offering. However the second and third could be based on thresholds (e.g. the probability of a win, and a derivation of predicted margins or profits).