

Improving the Invoice-to-Cash Process with Customer Payment Predictions

CSCI-P556: Applied Machine Learning

Fall 2019

Adam Hilgenkamp (ahilgenk@iu.edu)
Kaitlin Pet (kpet@iu.edu)
Jashjeet Singh Madan (jsmadan@iu.edu)
Surya Soni (susoni@iu.edu)

September 28, 2019

1 Problem Statement

According to Schwab (2016, p.1) “[The Fourth Industrial Revolution] is fundamentally changing the way we live, work, and relate to one another”. One trend of the Fourth Industrial Revolution that is highlighted by Klaus Schwab and the World Economic Forum is the increase of customer expectations. Customers increasingly expect a personalized and seamless experience, and businesses frequently rely on customer data to deliver that experience (Schwab, 2016, p.53).

An important part of interacting with a business is paying for the product or service. For businesses that send bills to their customers (instead of collecting payment immediately), this is known as the invoice-to-cash process. The invoice-to-cash process consists of invoicing the customer, collecting the payment from the customer and crediting the payment to their account. However, this payment is not always made on time for a variety of reasons. Some companies may not be able to pay and others want to hold onto cash reserves for as long as possible. Being able to predict whether a customer will pay their invoice on time has two main advantages for a business: a) helping the company effectively manage its cash flow expectations for its key stakeholders and b) improving the customer experience by reducing the need for intervention from collectors.

First, predicting whether a given customer will pay on time is important in optimizing customer experience in the invoice-to-cash process. Large businesses that sell products or services to other businesses typically invoice and collect payments from customers with some level of automation. Automating the process of creating and sending invoices is fairly straightforward, but the process of “following up” with customers is more labor intensive and time consuming. A one-size-fits-all approach to automating the follow up process would result in either flooding customers who usually pay on time with messages or not enough contact with customers who may not pay on time. With increasing customer expectations in mind, businesses need to find a way to avoid sending excessive messages to a customer who consistently pays on time while still being proactive to ensure payments are collected in a reasonable amount of time. Bucketing customers by their

predicted payment behaviors would allow a business to tailor "follow-up" strategies for each type of customer.

Another challenge for companies is forecasting the cash inflow in a given period and how close they would be to internal targets and external expectations from Wall Street. This forecast is typically calculated using general historical trends and qualitative information from the collections team. The ability to predict a) whether customers will pay their invoice and b) how late customers pay their invoice would give a company a better quantitative prediction on cash flow for a given period. The result of more accurate predictions allows the managers of the business to make proactive decisions such as paying an invoice early if they expect to collect more money in a period. The combination of greater accuracy and better customer experiences will be necessary for businesses to compete and succeed in the Fourth Industrial Revolution.

2 Data

In this paper we will be examining data from a Fortune 500 technology company. We have been given access to a redacted invoice dataset containing approximately 7.7 million observations and 139 features dating back to January of 2015. However, not all of the features are relevant or sufficiently populated so we will need to filter down our selection of features to reduce the noise in the data. To supplement the invoice data we have access to other datasets that we can use to add additional features. Additional datasets include account details which will allow us to bring in attributes such as the size of the customer (based on publicly reported annual revenue) and order details which will allow us to include the products purchased.

The dataset is redacted in order to remove any information that would identify a specific customer such as company name, address, or phone number. These fields are instead transformed into numeric IDs that obfuscate the fields but still allow us to use them in our models.

Each record in the invoice dataset is created when the invoice is sent to the customer. The record initially is filled with information about the invoice such as payment terms, payment type, billing country and the amount due. The invoice also has additional fields that are updated throughout the process. As part of our feature selection we will need to remove any of these fields that could "leak" information about the future to the model. For example, the last contacted date would be updated each time contact is made. This field would not be populated when a new invoice is observed by the model so the prediction using future information about the process will effect the error rate when the model is used to classify or predict real world data.

3 Research questions

When invoices are sent out to customers they are expected to pay by the due date. Late payments occur when customers either cannot pay or want to hold onto cash reserves and decide to pay late. The aim of this project is to predict the timing and outcome of the payment, i.e. informing business leaders if they can expect this payment within a certain time period. If the payment timing can be predicted with high accuracy, then the company can make decisions to manage its cash to meet stakeholder expectations and avoid intervention by collectors, hence increasing the overall customer satisfaction. Based on this goal, we have three research questions we would like to address:

1. Can we accurately predict whether a new customer will pay on time (classification)? This will help improve customer experience by bucketing customers into different communication strategies. For example, a 'low risk' customer may only receive one email while a 'high risk' customer would be personally called by a company representative.
2. Given information on if a customer is likely to pay or not how many days late will they pay (regression)? This will help the business better predict total cash inflow for a certain period.
3. Can we utilize case notes as an effective feature, i.e. can we use NLP techniques to automate analysis of notes instead of reading them on a case-by-case basis?

4 Evaluation Criteria

We will randomly split the dataset into training and test data. The training data will be split into several sections and models will be tested using cross-validation. Given the fact that our data points vastly outnumber potential features, we are unlikely to have issues with high variance. Therefore, inaccuracy in predictions will likely stem from high bias, i.e. we are more likely to underfit than overfit our data unless there are errors in the cross-validation process.

Work has also been done using Einstein Prediction builder, which is a Salesforce product that aims to democratize learning algorithms to the business user. It is a UI based AI which assists in building custom queries and formulae. We can compare performance of our model against the Einstein Prediction builder.

5 Timeline and Roles

Date	Deliverable	Research Question	Members Involved
October 11	Data cleansing done on base invoice dataset	General	Adam
October 12-18	Begin building models and testing on the base dataset	General	All
October 18	Join in additional fields to base invoice dataset	General	Adam
October 25	Finish building minimal working classification model	Question 1	Adam and Surya
October 25	Finish building minimal working linear regression model	Question 2	Jashjeet and Kaitlin
November 8	Test whether payment history (time series data) can be incorporated into regression model as predictor	Question 3	Kaitlin and Adam
November 8	Test whether sentiment analysis can be used to transform notes into predictive feature	Question 4	Surya and Jashjeet
November 24	Finish Building Models	General	All

References

Schwab, K. (2016). *The fourth industrial revolution*. World Economic Forum.