Customer Payment Prediction in Account Receivable

Hetul S. Shah

[1]Ahmedabad University, School of Engineering and Applied Science, Navrangpura, Ahmedabad, Gujarat 380009, India

Abstract: ***It is commonly agreed that AR (account receivable) is most valuable asset of any business firm. It can be a source of financial difficulties for firm when they are not efficiently managed and underperforming. So, it is important to identify data pattern in AR and get meaningful insight from AR data. This paper demonstrates how supervised machine learning can help to build model to predict payment outcome of invoices which are yet not paid (Open) based on historical data. Proposed method can predict with high accuracy that in which age bucket (i.e. On time,1-30,31-60,..,Over 150) the invoices will be paid. This method is implemented in the context of real-world AR data. Finally, simulation results are shown which proves that this approach can give high accuracy and save significant amount of collection time.***

Keywords: Predictive Modeling, Accounts Receivable, Supervised Classification, AdaBoosted Decision Trees

1. Introduction

We are entering in the era of big data in business analytics. Business analytics is divided in below three levels:

- Descriptive Analysis
- Predictive Analysis
- Prescriptive Analysis

Before, Machine learning was introduced business firm use to generate the visualized data which can be easily comprehended by humans and then they use to take business related decision manually and this is called descriptive analysis. This paper will propose the method for predictive analysis which is one step ahead of Descriptive analysis. In predictive analysis, we will try to predict the future trends using advance analytics algorithm (Machine learning). Prescriptive analytics is related to both descriptive and predictive analytics. Prescriptive analytics attempt to quantify the effect of future decisions in order to advise on possible outcomes before the decisions are actually made. This paper will illustrate how to implement predictive analysis on most valuable asset of business which is **Accounts Receivable**.

Accounts receivable are the lifeblood of a business's cash flow. Business's accounts receivable are important part of calculating business's profitability, and provide the clearest indicator of the business's income. The cut-throat nature today requires firm to utilize the data with full potential to survive and enable growth. Account Receivable represents the invoices which are yet not paid (Open) by customers. Account Receivable keeps track of invoices which are paid and their overdue days based on their due dates. Apart From that, it consists all the important information about customer and invoice. Typically, all big business firms generate thousands of entry monthly in their AR report. Machine learning empowers the ability to transform these abundance amount data into actionable knowledge.

In this paper, we will focus on improving the effectiveness of AR collections. We will use historical data of customer's paid invoices to identify the customer payment behavior and create the predictive model. This model will predict payment bucket of newly created invoices. This can help the collection to focus on delinquent invoices and take preemptive actions to drive down the collection time. In addition to that, we can prioritize the invoices based on the payment bucket, one can Optimize the use of collection resources.

2. Methodology

Invoice prediction task falls under supervised machine learning problem as we have historical payment data of customer and based on that we have to predict future payment outcome. We used **Adaboost decision tree** algorithm to accomplish this task. These problem falls under multiclass classification because there are 7 possible outcomes (On Time,1 – 30,31 – 60,61 – 90,91 – 120,121 – 150,Over 150).Basically, These age-buckets are the period in which invoice is paid after the due-date. So, If for any invoice if due date is 31^{st}, January and amount is paid on 10^{th}, February then age bucket of that invoice falls under '1-30'.

AdaBoost: Ada(Adaptive)boost(Boosting) is an iterative process that fits a sequence of weak learners on different weighted training data. It starts by predicting an original data set and gives equal weight to each observation. If the prediction is incorrect using the first learner, then it gives higher weight to observations which have been predicted incorrectly.

Like any supervised machine learning process, this process is divided in 5 stages,

1) **Raw Data**: In this stage, our model will fetch the date from data source to train and test the model. Real-life Historical data of Accounts Receivable was used as data source.
2) **Data Transforming:** According to researcher's Machine learning models are as good as the data that is used to train them. In this stage, Data transformation procedure took place to make the data ready for machine learning model. Firstly, there are some missing values in payment term

Licensed Under Creative Commons Attribution CC BY

which should be converted to zero and convert all the dates from string to date-time format. After this, Values of age bucket are quantified from 0-6.

3) **Feature Engineering:** This is most important stage to make most accurate predictive model to identify the features (parameter) which influence the outcome or prediction. Below listed features plays most important role to identify the payment trends,

Table 1: List of Features

Feature Level	No.	Feature Name
Invoice-Level	1	Invoice Amount
	2	Invoice Payment Term
	3	Invoice Due Month
	4	Gap b/w current and previous invoice due date
Customer-Level	5-11	Ratio Of Amount of Paid Invocies in each bucket by Total Amount of Paid Invoice for each Customer
	12-18	Ratio Of number of paid invoices in each bucket by Total number of Invoices for each Customer
	19	Average Amount of Paid Invoices
	20	Average Amount of Late Paid Invoices
	21	Average Due Period of all Invoices
	22	Average Gap b/w Paid and Late Invoices

4) **Training Model:** In this layer, the data with above listed 22 features is fed to Adaboost decision tree model. As training dataset, data except of past 2 month's data were taken to train the model and remaining data is used to test the trained model. We are using python's skylearn library to implement Adaboost decision tree multiclass classification.

5) **Prediction:** This is the last stage of process, where trained predictive model is tested on testing dataset which are past 2 months of data. In this stage, we will compare the prediction of past two months of data with original value to check the accuracy and have detailed analysis of prediction.

3. Results

For analyzing the accuracy of predictive model, We used skylearn's accuracy_score() function to calculate the result's accuracy and we got **85.2%** accuracy for around 27808 invoices which are generated in past two months. Apart from that, we generated the confusion matrix to get better idea of models accuracy for each individual age bucket. A confusion matrix is a table that is often used to **describe the performance of a classification model** on a set of test data for which the true values are known.

Table 2: Confusion Matrix

		Age Bucket Predicted						
		On Time	*1-30*	*31-60*	*61-90*	*91-120*	*121-150*	*Over 150*
Age Bucket Actual	OnTime	14158	767	101	45	19	18	10
	1-30	754	5326	202	48	21	10	7
	31-60	289	146	2006	121	33	20	9
	61-90	189	191	224	1444	104	31	11
	91-120	114	91	66	83	571	30	7
	121-150	53	56	32	12	28	232	4
	Over 150	1	6	5	8	11	33	61

As shown in table 2, Diagonal entries are higher compare to other values which clearly proves high accuracy of proposed model.

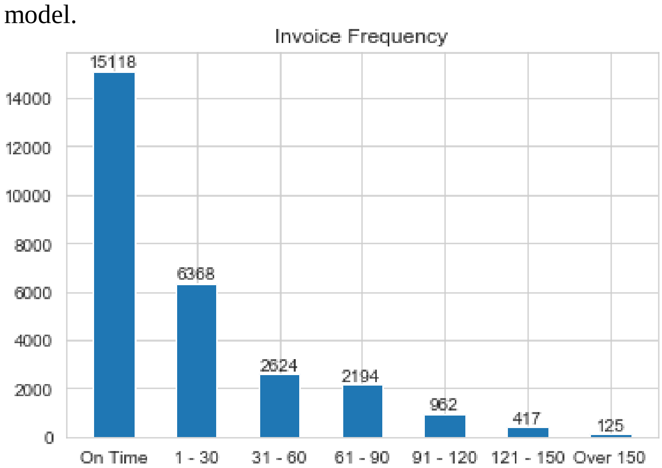

Figure 1: Number of Invoices in Each Age Bucket

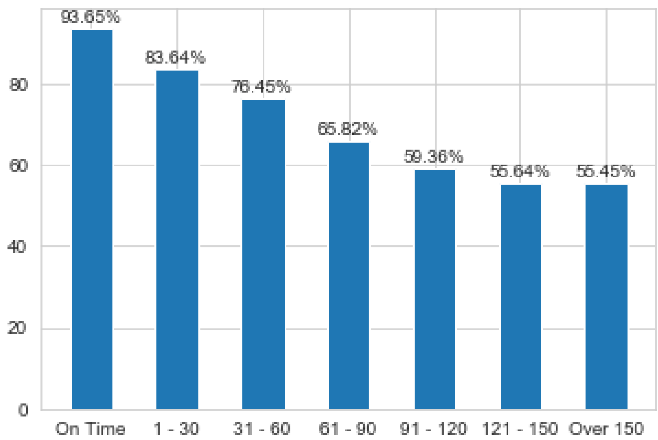

Figure 2: Accuracy of Invoices in Each Age Bucket

As shown in figure-1, Highest Number of Invoices are paid on time and then the count decreases as age bucket gets older (e.g. 1-30, 31-60, …, Over 150).

As shown in figure-2, proposed predictive model gets highest accuracy for age bucket On Time and then the accuracy decreases as age bucket gets older. Thus, average

Licensed Under Creative Commons Attribution CC BY

accuracy remains above 85%. It is difficult to get the exact accuracy for invoices with higher age buckets (i.e.121 – 150, Over 150) because they are extremely unpredictable and doesn't have any patterns. Plus, there are very few invoices which are paid in higher age buckets.

4. Conclusion

As popularity of machine learning is increasing in finance domain, this is encouraging researchers to identify the problems in finance domain which can be solved using machine learning techniques. This proposed method can help researchers to get one step closer to achieve full automation in most valuable asset of finance, which is Accounts Receivable. Machine learning is a natural fit to leverage the power of business analytics. Proposed method can help to understand the pattern of invoice and improve the account receivable collection by predicting the delay in advance. This method can also help big business firms to identify and prioritize the delinquent Invoices to save their collection teams time and resources. Proposed method gives more than 85% accuracy regarding invoice's payment outcome. So, it is safe to implement proposed method in real life to get good knowledge of future trends.

5. Future Scope

- Our first target is to improve the accuracy for higher age buckets.
- Implement prescriptive analysis to give feasible solution for late invoices.
- There are some invoices which are overdue for more than few months which are technically in dispute. So, identify the invoice which are likely to get into any sort of dispute.

References

[1] Sonali Nanda, "Proactive Collections Management: Using Artificial Intelligence to Predict Invoice Payment Dates," Credit Research Foundation 1Q 2018 Credit & Financial Management Review.
[2] Ji Zhu†‡ , Hui Zou§ , Saharon Rosset and Trevor Hastie," Multi-class AdaBoost", Statistics and Its Interface, Volume 2 (2009) 349–360
[3] "Descriptive, Predictive, and Prescriptive Analytics Explained" vnunet.com, para. 2, Jan. 31, 2001. [Online].Available: https://halobi.com/blog/descriptive-predictive-and-prescriptive-analytics-explained/
[4] http://scikit-learn.org/stable/modules/generated/sklearn.metrics.confusion_matrix.html
[5] John G. Salek, Accounts Receivable Management Best Practices, ISBN: 978-0-471-71654-9, July 2005.
[6] https://scikit-learn.org/stable/auto_examples/ensemble/plot_adaboost_multiclass.html
[7] Anup K Saha, Shaikh Masrick Hasan"Efficient Receivables Management A Case Study of Siemens Bangladesh Limited", Researchgate, January 2014.

Author Profile

Mr. Hetul S. Shah Obtained the B.Tech. Degree in Information and Communication Technology(ICT) from School of Engineering and Applied Science, Ahmedabad University at Ahmedabad, Gujarat. At Present, He is working at Stridely Solutions as Machine learning engineer.

Licensed Under Creative Commons Attribution CC BY