Discussion

* Evaluate the work critically
* Answer questions raised in the introduction
* Limitations
* Further research

Conclusion

* Summarize thesis
* No new information
* Reinforce main points and why are they important
* Clear understanding of the research and its contribution to the field

It has been the main objective of this work to assess the uncertainty which is associated with the PNBD model’s prediction of the individual customer lifetime values. Previous research on this topic was found to be scarce. To fill this gap, several methods that produce prediction intervals were identified in the literature and applied and adapted to the specific context of CLV prediction. Amongst others, a special focus was put on Conformal Prediction as a relatively young method that has recently gained importance in the statistical community. The resulting intervals were benchmarked with several metrics against an existing PI generating method from the R package CLVTools.

The central contribution of this thesis is the successfully implementation of methods that deliver intervals with high reliability on several real-world data sets, capturing the underlying uncertainty, which no other research has covered so far. It were especially the Bayesian Approach, Quantile Regression and Conformal Prediction which achieved a reliable coverage, which, however, comes at a cost. Capturing uncertainty adequately in the presence of high uncertainty causes wide intervals and little sharpness, which is the case for the PNBD model. Two other methods, including the bootstrap implementation from CLVTools, in contrast, create very short intervals, leading to low reliability but high sharpness. There is no method that provides a compromise between these two criteria.

A side objective was to assess the capacity of the constructed intervals to enhance the distinction of low- and high-value customers. Unfortunately, the PI generating methods hold mostly information about the uncertainty of the whole data set and then apply it to individual customers to form PIs. This lack of individual uncertainty information can be overcome by introducing a covariate that holds information about individual uncertainties. In this work, this covariate needed to be simulated but might exist for other data sets. With this additional information, the intervals were re-scaled and individualized and could eventually be used to help with the differentiation of customers.

Goal of this work:

* Benchmark bootstrap
* Alternative PIs
* Apply them in marketing

What methods were applied? (CP: Possible in this context)

Which results were achieved?

* First study to propose valid prediction intervals for CLV context
* Bootstrap was benchmarked
* Several methods applied and alternated
* Reliable PIs or sharp PIs
* No applicability for marketing

Limitations

* Only 4 data sets: Generalizability
* Speed of methods
* PIs do not hold individual uncertainty
* Not sure if such a covariate exists
* Conformal prediction and Quantile Regression need past data to work
* No method that offers a compromise

In the course of this work, it was no method found that achieves both, a high reliability and sharpness at the same time and none of the used methods incorporates directly individual uncertainty. In addition, quantile regression and conformal prediction need past data to work, what might pose a major obstacle in real-world scenarios. Regarding the application in identifying especially valuable customers, it was necessary to simulate a covariate, as the data sets used did not provide a suitable one. If such a covariate actually exists in practice is not certain but identifying one and verifying the results from this work in reality could be subject to future research. Also, the improvement of PI generating methods, in terms of e.g. incorporating directly individual uncertainty, improving speed, especially for the Bayesian approach, or developing a method that offers a good compromise between reliability and sharpness are areas that need further attention. This work utilized four datasets to ensure the generalizability of the results. However, confirming the findings with additional datasets from different industries and exhibiting diverse customer behavior would be beneficial.

Future research

* Develop methods that internally incorporate individual customer uncertainty, not just their CET
* Methods that offer a compromise
* Confirm results with other data sets and real-world covariate
* Faster methods