**Capstone Project overview**

Perrine Mignot

Data Science and AI Course student - Part-time course ending 3/12/2020

1. Project overview

A home loan provider has approached our marketing agency to analyse their conversions in Australia. They would like some insights to understand what are the key factors driving customers to submit an application on their site from a marketing perspective. The goal is to help them prioritise resources, time and money in the right areas to improve conversion rate and spend less budget in Marketing.

1. Problem statement and question

The company is spending millions of dollars every year in Marketing in various areas to attract and convert visitors and would like to be smarter about where they spend their budget. Analysing the conversions and what drives them will give them some great insights into which specific aspect is making customers submit an application.

* Business question: What are the key factors leading to conversion?
* Data question: Can we predict customer’s conversion and identify the features impacting conversion?

The desire outcome would be to define a few key areas where the company can focus on to reduce expenditures and increase conversion. An added bonus will be to obtain a better understanding on their prospects/customers behaviours.

The company hasn’t had the chance or time to research deeply into their own conversions yet. Data science and even analysis is a fairly new area in Marketing and it requires some preparation and setup as every behaviour needs to be tracked. There is a lot of information about customers’ behaviour online but nothing specific to their industry and situation. We see more and more companies willing to get a better understanding of their own data and what insights it can bring to become more efficient with their resources. This company is one of them.

1. Industry overview

This company is a small home loan provider operating globally but interested in their Australian market, we cannot name them for privacy reasons.

They provide home loans at competitive rates and their objective is to reduce the length of home loans for Australians.

There are more and more smaller home loan providers and financial institutions in the market and competition is fierce, especially online.

1. Stakeholders

The stakeholders are the Marketing team at the home loan company as well as the senior management team at the Marketing agency who are across the Marketing activities and expenses and interested to see how these insights can help them be more efficient and reduce costs.

1. Data

We have access to the company’s data and have organised behavioural tracking, the goal is not only to analyse user behaviours online which can be done simply, but behaviours associated all the way to conversion.

The data includes 155.000 rows which represent visits to the site and 39 features. The features include information about the visitors, their behaviour on site and if they converted or not.

The data is highly reliable and the quality is very good.

We cannot share the dataset for privacy reason.

1. Data analysis

The key highlight of the analysis was the number of days a visitor takes to submit an application.

We also discovered some fantastic insights into customers location, visitors behaviour on site and which Marketing activities have a good conversion rate.

1. Modelling

We used h2o has most features were categorical and sent most features to the model, except the ones highly correlated with conversion like loan type and number of applications submitted as they produced leakage.

We also had to randomly under sample the data as it was heavily imbalanced (conversion rate is 0.8%). We created 10 tables in pandas that included the same 700 conversions first and we then appended 10 different sample tables made of 700 random visitors who didn’t convert. The goal was to have a few small balanced datasets. We did this manipulation 10 times as our datasets were therefore small and we needed to ensure the stability of the model, so we compared them all and made sure they all had similar and strong results.

We tried Gradient Boosting but it wasn’t stable as it returned a perfect model. We then tried random forest which performed very well so we used random forest to check our 10 tables. The models were fast and results strong every time.

We needed to predict the conversions to then find out what factors led them to convert. The scores we looked at where accuracy and auc for the overall performance but we specifically looked at recall and precision to focus on the positive predictions. We want to make sure the errors around positive predictions are minimal, especially precision since we want to reduce False positive as it weakens our results.

1. Outcomes

We found that time of visit and time to conversion were key components affecting conversions. This is valuable insights as it helps guide the marketing agency and company to target customers at specific time only and reduce expenses.

It would be interesting to analyse other subsequent features further like Marketing source and obtain the money spent in various Marketing channels to understand the conversion rate based on the money spent.

We could also go further into our model and do stacking ensembles to get an even better performance.

The business and data question were answered. We are confident with the data, analysis and model performance and we believe the insight will be valuable to reduce Marketing expenditures and increase efficiency.

The model can be run again in the future with more data and perhaps features to get even more insights.

1. References

The analysis, presentations and appendix are available at [github.com/PerrineLeung/IoD\_Assignments/Capstone project](https://github.com/PerrineLeung/IoD_Assignments/tree/master/Capstone%20project)