Assignment 3 - Predicting Companies' Fast Sales Growth

A. BOGNAR

Introduction

Our investment firm is looking for businesses with potential for rapid growth to invest in. To this end, prediction models were built to find fast growing firms.

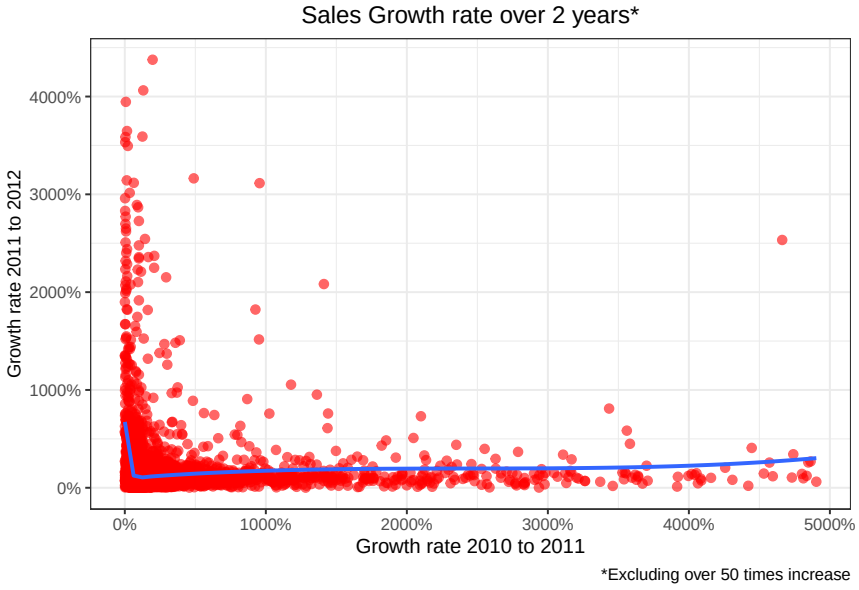
Below I present the main findings. All supplementary code is available on Github.

Prediction

Several options were considered to measure company growth. Relative change of yearly sales was chosen as the target variable because:

- Unlike Profit and other measures, Sales was available for high number of firms across multiple years.
- A rapid, but not unrealistic, increase in company sales is associated with lasting growth.

The below graph shows the yearly sales growth rates of companies in the 2 years leading up to our analysis. Negative or very small as well as extreme (above 50 fold) increase is negatively associated with further growth next year, but realistic and even huge sales increases tends to lead to further growth in proceeding years. Note that the below graph excluded over 50 fold increases where a slight negative association is seen, there are few such companies.



Therefore, our target became to predict which companies will increase their sales by at least 15 percent from 2012 to 2013 which is still a relatively high bar.

Our predictors included:

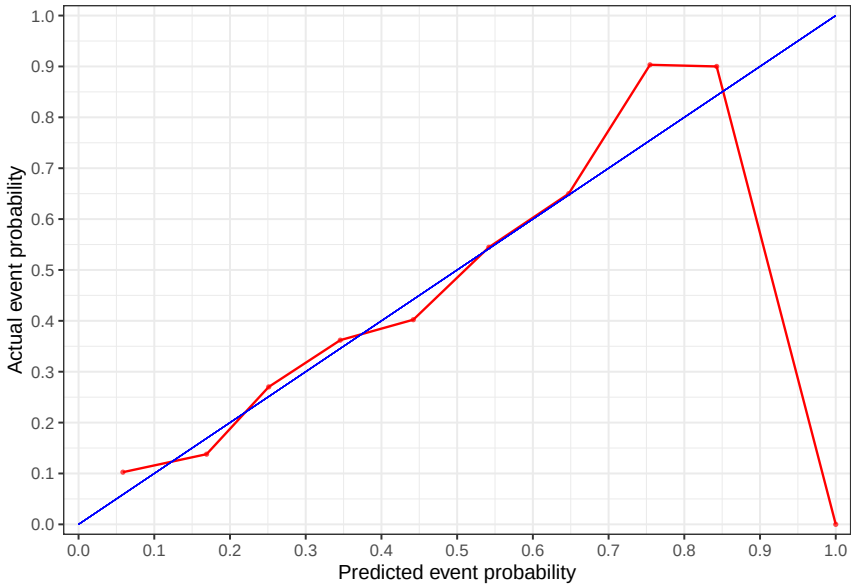
- Past two years' sales.
- Figures from financial reports such as invetories.
- General company characteristics such as age.

Models, Results

Both theory driven - logistic - and data driven - Lasso and Random Forest - models were used. Below you can see the initial results. Of the logit models, I will describe the findings of Model 4 as it has one of the lowest errors and most of its predictors seem to be relevant for the prediction.

	Number.of.predictors	CV.RMSE	CV.AUC
X1	7	0.4600	0.5734
X2	14	0.4559	0.6053
X3	30	0.4518	0.6334
X4	74	0.4493	0.6419
X5	87	0.4491	0.6447
LASSO	57	0.4495	0.6193

Below you see Model 4's performance based on how confident each prediction is. For the most part it performs well but for very confident predictions we see a sharp decline. That is because there are few companies confidently predicted to rapidly grow. In fact, there is only one company the model predicted with above 90% to grow which turned out to decline which is the error shown. This is a very unique company and while it's sales declined from 2012 to 2013, the year after that it increased 47 fold.



Evaluation

The business decision to make based on the prediction is whether to invest in a company in hope of growth or not. We are looking to make investments around 100 thousand USD to jumpstart companies. Our baseline investments make around 10% interest yearly, thus losing a potential company with 15% increase will cost us 5000 USD. On the other hand, we estimate that investing in companies that underperform expectations will on avg. cost triple that amount, 15000 USD. With these assumptions Model 4 produces the below decisions. As the loss of investing in declining firms is high, very few predictions for growth are made (42) but the majority of them accurate. This is promising as after model selection, the companies would individually be evaluated likely to produce good success rate.

	no_fast_growth	fast_growth
no_fast_growth	2585	1176
fast_growth	5	37

On the other hand, the conservative approach will likely miss many potential growing companies. Based on budget constraints a more risk tolerating model can be built to find additional companies.

Finally, below are the main figures of the 3 types of models built The data driven Random Forest somewhat outperforms the other two. All models produce high bars for predicting company growth and the avg. expected losses on the predictions is very close at around 1500 $ for each model.

	Number.of.predictors	CV.RMSE	CV.AUC	CV.threshold	CV.expected.Loss
Logit X4	74	0.4493	0.6419	0.7066	1539
Logit LASSO	57	0.4495	0.6193	0.6887	1549
prob Random Forest	38	0.4438	0.6685	0.7141	1522

Limitations, Further Analysis

As seen from Model 4's results, unique, extremely growing/declining companies are difficult to predict. For more accurate results a more similar set of companies based on range of sales, industry may be investigated.

Additionally, any one company may experience unexpected changes such as reconstruction, mergers etc. That is why individual inspections are still necessary prior to investments.

As mentioned based on business requirements lower bar for predictions may be set to include more potential companies at the expense of increasing the risk of underperformance.