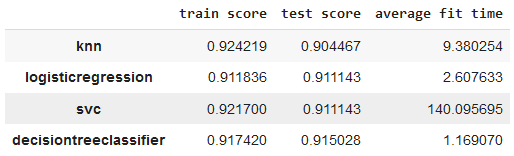
Link to notebook

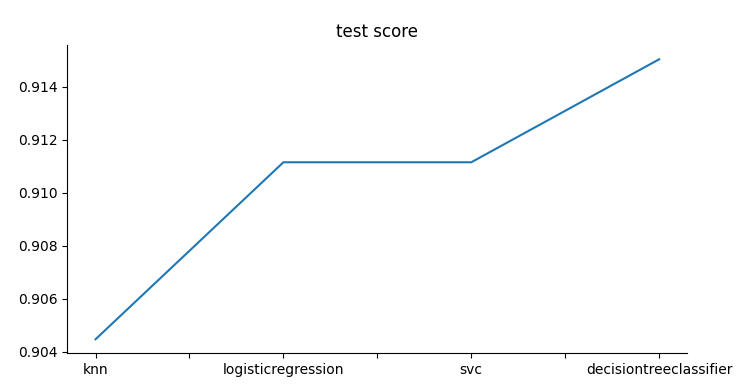
<https://github.com/changliu8888/bank-telemarketing>

## Summary of findings

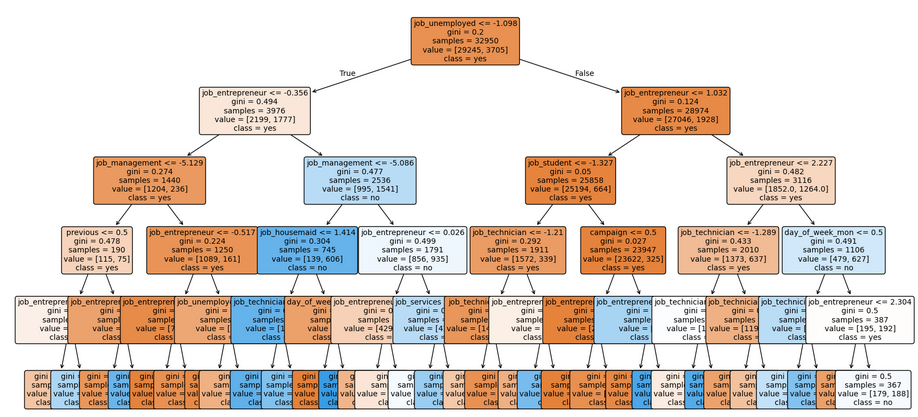
* The rate of “yes” response is highest when the bond rate is low, and when total employment is low.
* The conversion rate is higher when consumer price index is low and when consumer confidence index is low.
* Anyone who has had earlier successful conversion will have a much higher chance of saying yes.

As for the fitting strategies, it is found that the decision tree classifier is the most accurate and economical.





It also produced the most insight and understanding of the classification.



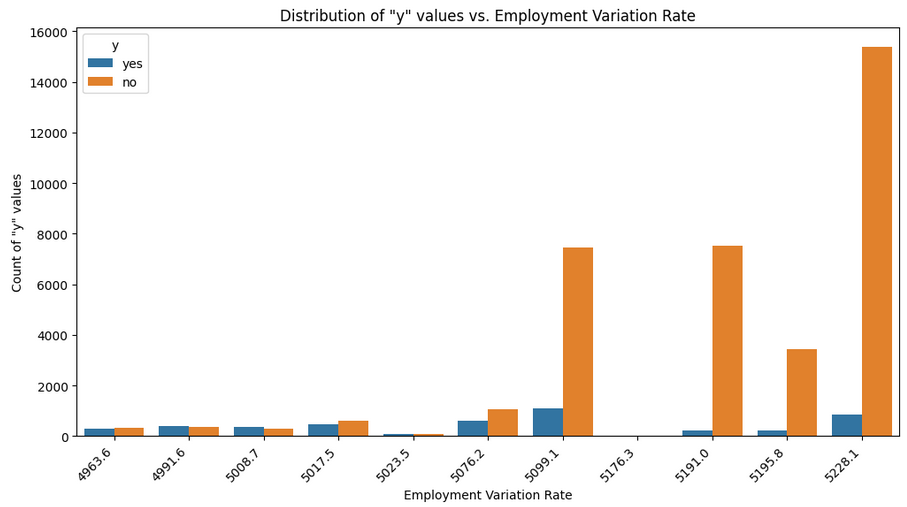
## Database explanation

"emp.var.rate" stands for "employment variation rate", "cons.price.idx" represents the "consumer price index", "cons.conf.idx" is the "consumer confidence index", "euribor3m" refers to the "3-month Euribor rate", and "nr.employed" means "number of employed people" - essentially a set of economic indicators used to assess the overall economic climate when analyzing factors like bank loan applications or consumer spending.

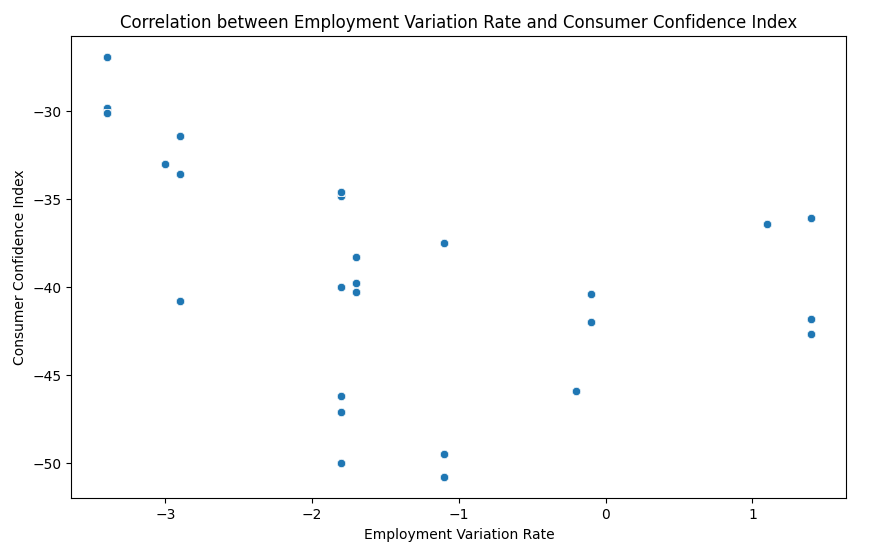
There are four datasets: 1) bank-additional-full.csv with all examples (41188) and 20 inputs, ordered by date (from May 2008 to November 2010), very close to the data analyzed in [Moro et al., 2014] 2) bank-additional.csv with 10% of the examples (4119), randomly selected from 1), and 20 inputs. 3) bank-full.csv with all examples and 17 inputs, ordered by date (older version of this dataset with less inputs). 4) bank.csv with 10% of the examples and 17 inputs, randomly selected from 3 (older version of this dataset with less inputs). The smallest datasets are provided to test more computationally demanding machine learning algorithms (e.g., SVM).

## Understanding the data and findings

It is easy to see the trend of response as a function of the consumer confidence, the bond rate, and overall job market.



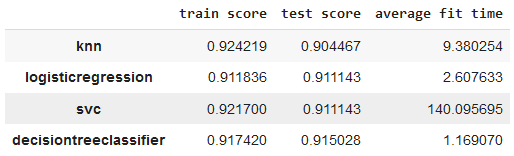
There is overall negative relation between consumer confidence index and employment variation rate.



## Modeling

Multiple models are used and cross-validated. The time, train score, test score are compared (shown below).

It is found that the decision model model gives one of the best results and still manage to use a fraction of time compared to SVM model.



## Business case understanding

1. The best way to design the telemarketing campaign is to do it when the interest rate is low, consumer price index is low (no inflation), and when overall economy is not the best. Many features such as education, income, marital status does not have material correlation to the final decision to accept telemarketing offer or not.
2. The best way to analyze the data is to use a decision tree model.
3. The people who responded earlier have a high chance of responding again. This seem to suggest the product is good.

## Next step and recommendation

It is so far not clear the year and the overall income. The relations of response to these critical parameters may be important. Future data sets could include such information.