Case study of the "Risk Assessment GmbH"

You are the new extraordinary Data Scientist of the "Risk Assessment GmbH" and you really want to become a part of this company. The "Risk Assessment GmbH" is a leading company for modelling specific risks of customers. They develop models based on different statistical and machine learning algorithms. As part of your job you have to solve a specific case study for this company and provide a solution which could be implemented. Remark: There are several ways to solve this task, but please find a reasonable way!

The Risk Assement GmbH (RA) wants to compete with SCHUFA which is the the biggest German credit bureau supported by many creditors. The aim of RA is to find a credit default risk model which fits to their customer data and which is able to predict credit default.

- Consider the *RAcredit_train.csv* as the data set, that you should analyse.
- For your model predictions, please use *RAcredit_test.csv*.
- **Hint:** `l_state` is the status of a loan which you must predict.
- `Default` is default and the `Fully Paid` is our desired state.

1. Import the data in R and prepare the data to use it. Your data set contains specific customer features and their corresponding loan state.
2. Do you face any problem and how would you solve them?
3. Define a suitable model which you fit to the data.
4. Evaluate the out-of-sample performance of your model and provide a comparison to at least one other modelling approach.
5. Identify all possible variables which are beneficial for the model.
6. What are the assumptions and limitations of your model?
7. This model should be an easy implementation, if you would have a higher budget and more time, what could you provide in addition to this approach?

Remark: The **R** code is very important for this task. Please spend enough time to provide a satisfying **R** program. Moreover, the data set has to be prepared and analysed. Furthermore, comment and interpret your results. The RA GmbH should be convinced, that you are able to obtain great results! Note; To compare the results among all applicants, we require an **R** program. If you use python, please explain why you done so?