

Fontys Hogescholen ICT

# Project Proposal

Loan Prediction by Zhaklin Yanakieva

Zhaklin Yanakieva

Eindhoven, [08/03/2021](#)

<b>Project definition</b>	<b>2</b>
Background	2
Project Goal	2
Project Domain	2
Domain research	3
Stakeholders	3
<b>Planning</b>	<b>4</b>
<b>Ethical considerations</b>	<b>5</b>
<b>Impact assessment</b>	<b>6</b>
<b>Exploratory Data Analysis</b>	<b>7</b>
<b>Conclusion</b>	<b>7</b>
<b>Bibliography</b>	<b>7</b>

# Project definition

## Background

The recent significant increase in loans has generated interest in understanding the key factors predicting the non-performance of these loans.<sup>1</sup> The idea behind this ML project is to build a model that will classify how much loan the user can take, depending on certain data that will be required by the person, signing for a loan, such as name/education/marital status/number of dependents, employment, etc.

## Project Goal

This project is going to be a prototype of a real loan predictor, which goal is to predict if a certain person is suitable for a loan.

## Project Domain

The domain of this project is determined to be: *Prediction of a loan*

Also the following research methods were applied:

- Literature study
- Stakeholder analysis

## Domain research

## Literature study

### What is required to receive a loan?<sup>2</sup>

---

<sup>1</sup>

[https://www.researchgate.net/publication/318816798\\_Project\\_Report\\_Student\\_Loan\\_Repayment\\_Prediction](https://www.researchgate.net/publication/318816798_Project_Report_Student_Loan_Repayment_Prediction)

<sup>2</sup> <https://www.forafinancial.com/blog/working-capital/8-bank-loan-requirements/>

## Are You Aware of These Bank Loan Requirements?

1. Purpose of Loan
  2. Business Experience
  3. Business Plan
  4. Credit History
  5. Personal Information
  6. Financial Statements
  7. Collateral
  8. Cash Flow
- ### How to Qualify for a Personal Loan

<sup>3</sup>There are many steps to take to qualify for a personal loan, with the first being to make sure that it's right for you. For example, if you want to borrow money to remodel your house or buy a car, a home equity loan or an auto loan may come with a lower interest rate. Unlike unsecured personal loans based solely on your creditworthiness, these loans are secured by the home you want to fix up or the car you want to buy.

Although paying for a family vacation or consolidating debt fits into the personal loan category, you may also want to check into a 0% introductory APR credit card. If you go that route, however, be sure that you can pay off the balance before the 0% rate expires.

## What makes people have the desire/need to have a loan?

<sup>4</sup>Personal loans are borrowed money that can be used for large purchases, debt consolidation, emergency expenses and much more. These loans are paid back in monthly installments over the course of typically two to six years, but it can take longer depending on your circumstances and how diligent you are with making payments.

Here are the top nine reasons to get a personal loan and when they make sense:

1. Debt consolidation.
2. Alternative to a payday loan.

---

<sup>3</sup> <https://www.investopedia.com/articles/personal-finance/010516/how-apply-personal-loan.asp>

<sup>4</sup> <https://www.bankrate.com/loans/personal-loans/top-reasons-to-apply-for-personal-loan/>

3. Home remodeling.
4. Moving costs.
5. Emergency expenses.
6. Appliance purchases.
7. Vehicle financing.
8. Wedding expenses.
9. Vacation costs.

## Stakeholders

Name	Details
Students	People who need a loan in order to finish their education and also are in need of a maintenance loan for a living.
Labour force	Working people who either need a loan for new purchases, or for living costs.
Self-employed	Working for oneself as a freelance or the owner of a business rather than for an employer.
Banks	Institutions which decide whether a person can be approved for a loan or not.

## Planning

In this chapter, I am going to explain the project planning for the upcoming weeks this semester and how I am going to achieve my goal. It gives me a better understanding of the

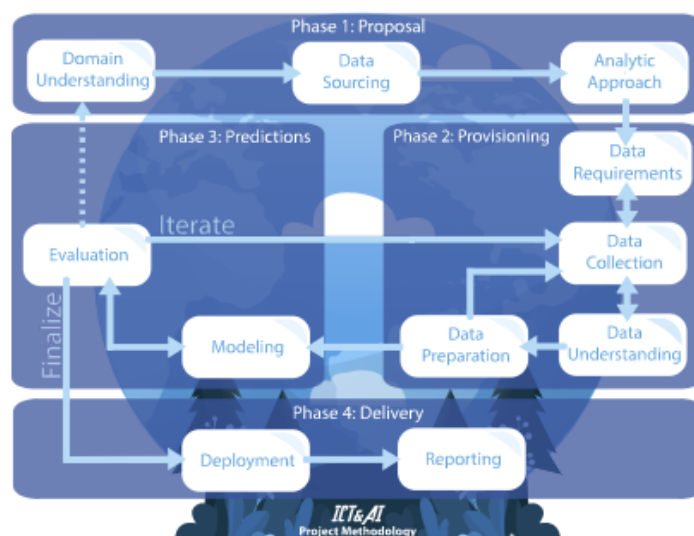
structure of the semester and also helps me to deal with deadlines. I am using my own sprints that are represented in the table (see Picture 1).

Part	Weeks	Project days	Phase(s)
A	3-6	12	1 and 2
B	7-10	10	2 and 3
C	11-12	10	4

(Picture 1:Phases)

From week 3 until week 6, this is the first phase (phase 1) of the project. This is where the business proposal and the exploratory data analysis report must be delivered, however I also decided to start with the first steps of the provisioning phase.

The next phase is week 7 until week 10. That is the time for the second part of the project (phase 2 and 3). The reason for doing phase 2 and 3 in the same sprint is because in the second iterations, the product has to be improved compared to the first one. This is the same for the final part of the project (phase 2,3 and 4). That is the final sprint where I have the final product ready.



(Picture 2: Phases of the project)

As you can see on Picture 3, the 4 weeks will be spent on the Project Proposal, the EDA ( Exploratory Data Analysis), Data requirement and Data collection ledger. After the 6th week I am going to spend the time on the Provisioning and Prediction phases. At week 10 to 12 , there will be the deployment of the project presented.

Step	Weeks	Phase	Deliverables
1	3 – 6	Proposal	Project Proposal, Exploratory Data Analysis Report, Data requirement, Data collection ledger
2	7 – 10	Provisioning	Data Analysis Report, Preparation notebook
3	14 – 15	Prediction, Deployment	Modelling notebook attached, Deployment report

(Picture 3: Planning of the project)

## Ethical considerations

In this part of the project, we are going to use the “The Rights Approach”<sup>5</sup> – a method which is focused on the individual's right to choose for herself or himself. It is a violation of human dignity to use people in ways they do not freely choose, so what should be taken into account is: the right of privacy and the right to what is agreed. In these cases, applicants have the right to present certain information only to the institutions they are asking for a loan.

### Ethical dilemmas:<sup>6</sup>

- **Is the development of “loan predictor” ethical?**

I consider the development of this project to be ethical because it benefits people without harming anyone. Also different levels of consent required from the people assigning for a loan can be enforced.

---

<sup>5</sup> <https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/thinking-ethically/>

<sup>6</sup> <https://www.nature.com/articles/s41599-020-0501-9>

- **What are the actions the involved stakeholders are carrying out to address the ethical concerns?**

I assume that part of the stakeholders - students, labour force, self-employed, will not have to take the ethical concerns into account, however, the institutions that approve them for a loan will be responsible for this.

- **What consent will be processed?**

The data that will be required so as the predictor to be manufactured will be based on the name, loan data, age, etc. In this case, the data will be processed in a way to create a result if the person is suitable or not for a loan.

- **Is there a violation of moral rights involved in the development of the “loan predictor”?**

As soon as the privacy of the client is met, which means that only the person that is applying for a loan and the bank that gives it have the right to see the data, there is no violation.

## Impact assessment

**NAME: Loan Predictor**

**DATE: March 8, 2021 12:46 AM**

### **DESCRIPTION OF TECHNOLOGY**

The recent significant increase in loans has generated interest in understanding the key factors predicting the non-performance of these loans. The idea behind this ML project is to build a model that will classify how much loan the user can take, depending on certain data that will be required by the person, signing for a loan, such as name/education/marital status/number of dependents, employment, etc.

### **IMPACT ON SOCIETY**



In their lives, at least once, people had the opportunity to have a loan. The prediction project will help lenders (banks) and people who want a loan, know if they are suitable for a loan. It will have a more beneficial than harmful effect to society. There can be some stigmatising effects for people who are in need for a loan, but do not cover the requirements.

## **HATEFUL AND CRIMINAL ACTORS**

The data that will be processed is going to be available only to the institutions that give out the loans, so in the case of creating this project, no one's privacy will be hurt.

## **PRIVACY**

The technology should take the name/education/marital status/number of dependents, employment. Taking into account these labels, accurate calculations will be made to determine the loan status of a person.

## **HUMAN VALUES**

The technology is rather beneficial to the users than harmful. It does happen to appear certain drawbacks for people who are unemployed or self-employed because it is usually more difficult to determine their loan status according to the data they apply. However, it is not impossible for them to be approved for a loan, so the project will be as beneficial for them as it is for the other people (students, labour force).

## **STAKEHOLDERS**

- Students
- Labour force
- Self-employed
- Banks

## **DATA**

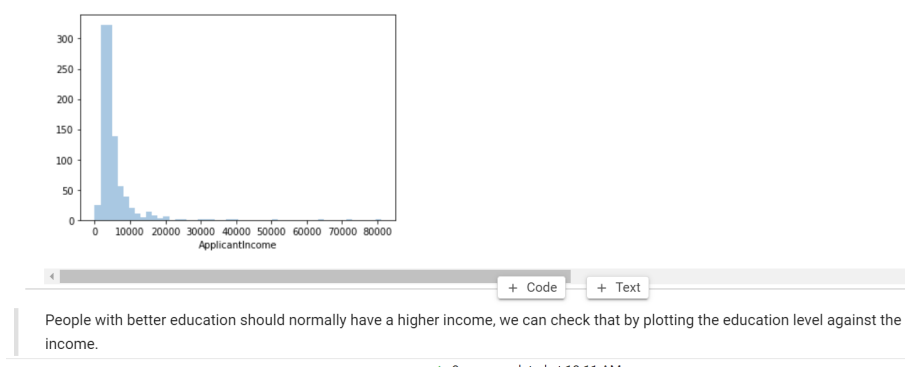
Depending on the available data, the research will be made, but there are always limits. This is clear to me, so the technology will be made with the awareness of these limitations and the users will be notified about them before beginning their experience with the application.

## INCLUSIVITY

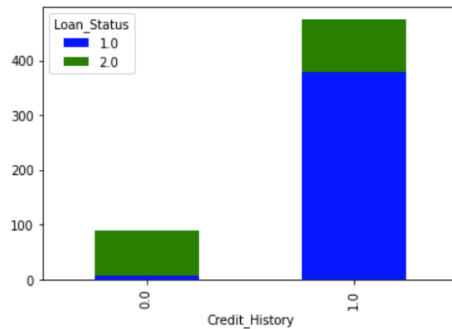
Due to political issues or historical issues, the banks are biased and therefore, there is some bias in the built-in of the Loan predictor.

# Exploratory Data Analysis

In the EDA, I separate the analysis into several parts. The first part is the Data collection. For this, I decided to use extraction of the data from a csv file as a technique because it gives the opportunity to work with two data sets that are given since I cannot use any back information because of its confidentiality. Then, the following part is the Data preprocessing, for which I tried different types of data transforms to expose the data structure better, so the model accuracy may be improved later.



> Here I am exploring the distribution of the numerical variables mainly the Applicant income and the Loan amount. What can be noticed are quite a few outliers.



This shows that the chances of getting a loan are higher if the applicant has a valid credit history.

You can find the exploratory data for test analysis [here](#).

## Modelling

The modelling stage will consist of training a Logistic Regression and Random Forest model. The reason for choosing to train this machine learning model is based on the explored information - the predicting label, which is a numeric value for which the models are suitable. In this case, the predicting label is the price of a real estate. The loan status is displayed as a yes-or-no answer.

## Evaluation

This chapter explains how I will assess the performance of the project and what features will be optimized.

After exploring the data set and training the models, a vital part of the project is the evaluation of the models performance.

First of all, the performance of the model has to be tested with common evaluation metrics. Then, an approach, which consists of predicting a future value, to evaluate the model is accurate enough to determine the performance.

# Deployment

Ideally the model will be deployed by using an API and creating a Web application.

The users will have the chance to access the Web application so as to query the API, which will be open for it.

# Conclusion

The emergence of this technology has risks and ethical concerns if it is managed properly it will lead to better quality of life for human society. Be that as it may, in the event that we hit the nail on the head, we will release the full advantage of AI for mankind. The proposal demonstrates that we ought to not really believe a machine learning algorithm. It is obviously not moral to utilize a machine to make decisions when we don't confide in it for using good ones.

Taking into account the data set, a couple of models were used in order to calculate the accuracy in the most efficient way. From the research, it could be concluded that the prototype of the project is feasible.

# Bibliography

*Forafinancial*, 26th July 2019,

<https://www.forafinancial.com/blog/working-capital/8-bank-loan-requirements/>.

*Investopedia*, 22nd January,

<https://www.investopedia.com/articles/personal-finance/010516/how-apply-personal-loan.asp>.

*Bankrate*, 11th January 2021,

<https://www.bankrate.com/loans/personal-loans/top-reasons-to-apply-for-personal-loan/>.

*Research gate*, July 2017,

[https://www.researchgate.net/publication/318816798\\_Project\\_Report\\_Student\\_Loan\\_Repayment\\_Prediction](https://www.researchgate.net/publication/318816798_Project_Report_Student_Loan_Repayment_Prediction).

*Edu*, August 2015,

<https://www.scu.edu/ethics/ethics-resources/ethical-decision-making/thinking-ethically/>. Accessed March 2021.

*Nature*, 17th June 2017, <https://www.nature.com/articles/s41599-020-0501-9>. Accessed March 2021.

