

Congratulations on finishing the Linear Algebra course! It is an essential part of your training as linear algebra is the mathematical foundation for many Machine Learning algorithms.

To test what you've learned, we have a project for you that demonstrates a practical application of linear algebra combined with its mathematical beauty. When you finish, send your work to the project reviewer. You will receive feedback within 24 hours. After that, you will make any necessary changes to your work and send it for a second review.

As you know by now, this process is often repeated several times until you get the green light from the reviewer and all the corrections have been approved. Your project will be considered complete once the project reviewer approves it.

The Sure Tomorrow insurance company wants to solve several tasks with the help of Machine Learning, and you are asked to evaluate that possibility.

- Task 1: Find customers who are similar to a given customer. This will help the company's agents with marketing.
- Task 2: Predict whether a new customer is likely to receive an insurance benefit. Can a prediction model do better than a dummy model?
- Task 3: Predict the number of insurance benefits a new customer is likely to receive using a linear regression model.
- Task 4: Protect clients' personal data without breaking the model from the previous task.

It's necessary to develop a data transformation algorithm that would make it hard to recover personal information if the data fell into the wrong hands. This is called *data masking*, or *data obfuscation*. But the data should be protected in such a way that the quality of machine learning models doesn't suffer. You don't need to pick the best model, just prove that the algorithm works correctly.

Project Instructions

1. Load the data.
2. Check that the data is free of issues — there is no missing data, extreme values, and so on.
3. Work on each task and answer the questions posed in the project template.
4. Draw conclusions based on your experience working on the project.

There is some precode in the project template, feel free to use it, some precode needs to be finished first. Also, there are two appendices in the project template with useful information.

Data Description

The dataset is stored in file `/datasets/insurance_us.csv`. You can download the dataset here.

- **Features:** insured person's gender, age, salary, and number of family members.
- **Target:** number of insurance benefits received by an insured person over the last five years.

Project evaluation

We've put together the evaluation criteria for the project. Read this carefully before moving on to the task.

Here's what the reviewers will look at when reviewing your project:

- Have you followed all the steps of the instructions?
- Have you kept to the project structure?
- Have you kept the code neat?
- Have you developed all necessary procedures and answered all the questions?
- Have you made your conclusions?

A more detailed checklist can be found in the end of the project template. You have your takeaway sheets and chapter summaries, so you are ready to proceed to the project.

Good luck!