Call for tenders for the creation of a cost prediction model

Project Description

The goal of the project is to quickly predict costs and times using artificial intelligence. To achieve this, we plan to create a demonstrator (a mockup), which is the subject of this call for tenders.

The end customers are companies that prepare estimates in all fields of activity. They collect information on the estimates they produce over the years and build databases containing various parameters that they want to use to predict the costs of new projects and thus avoid having to realize detailed estimates.

We want to create a mockup that would demonstrate the effectiveness of the proposed method. The goal is to train a model that could make predictions with a low margin of error.

We will prioritize the use of Copilot and ChatGPT because our software is developed primarily with Microsoft tools, but we are of course open to any more powerful tool when it is sustainable and inexpensive. The choice of product(s) should be mentioned in your response.

To build this model, we were able to collect data from numerous estimates (around 3,500) produced between 2008 and 2025 in France. The data provided has been anonymized and adapted to avoid any correlation with real projects. To simplify exploitation, all amounts are expressed in euros, all labor times are expressed in hours, and all vehicle times are expressed in days. The Excel file provided contains two sheets containing the same data, one in English and the other in French. You can use the language that suits you best.

The data provided comes from companies working in the electricity field, both tertiary and industrial, and, for simplicity, we have extracted estimates from “Power and Lighting” parts only.

There are two main project types: TERTIARY and INDUSTRIAL (in column B) and project sub-types called "Type of project" (in column C): DATA CENTERS, OFFICE BUILDINGS, LOGISTICS PLATFORM, CAR PARKS and MEDICAL BUILDINGS. Of course, depending on the project type, the relationships between the different columns are different. For example: it takes longer to setup electrical cables in a car park environment than in an office environment.

To estimate a new project, the end user will have to enter some mandatory input data:

1. Project type (column B)
2. Kind of activity (column C)
3. Building area in square meters (column D)
4. Project year (column E)

Costs are highly dependent on the building's surface area, of course. The project year also has a significant influence because it reflects inflation. Finally, the kind of activity massively influences

Written by : Pierrick Combreau
Date : May, 7 2025

construction times. All these correlations must be "found and suggested" by the artificial intelligence engine to enable the model to be trained.

The four pieces of information listed above should be used to provide estimated costs, i.e. to estimate what is found in columns AX, AY and BA. The content of column AZ (margin) results from a calculation between the total cost (column AY) and the sales price (column BA).

The user will be able to add technical input data that he could find in the tender documents such as:

1. Number of transformers (column I)
2. Transformer power (column H)
3. Number of low-voltage general panels (column N)
4. Number of low-voltage auxiliary panels (column Q)
5. Number of inverters (column T)
6. Number of capacitors (column W)
7. Length of cable trays (column Y)
8. Number of light fixtures (column AA)
9. Number of switches (column AC)
10. Number of sockets (columns AD and AE)
11. Number of emergency fixtures (column AG)

Based on this input data, the model will look for correlations and predict the other columns. The software can be asked to provide examples close to the entered characteristics. These examples can come from previous years because the model will take the inflation into account to provide the new values.

Interface

Our teams work exclusively with Visual Studio, in C# with the .NET Framework or .NET Core. It's essential that the delivered code be easily usable by our teams and integrated into our software. Our ultimate goal, after creating the demo mockup, is to offer our customers an interface that can be adapted to their data. This data, of course, is completely different for a business field other than electricity.

If a database is required, it will be Microsoft SQL Server. Generally, all components will preferably come from Microsoft.

However, the aim of this call for tenders is simply to produce a demonstrator for the electrical sector powered by the data provided in the attached Excel file.

We will therefore favor a Windows Forms application written in C# with an interface allowing the user to enter the input data and train the model. However, we could also accept an application written in VBA under Excel if you do not have the skills to write in C# (you will need to specify this in your response to our call for tenders)

It is important that the end user can know how the proposed results were evaluated, in particular by producing a list of existing projects close to the project to be carried out so that he can see if the data seems consistent to him.

At launch, a window would be displayed with all the criteria listed. The user would have to enter at least:

1. Project type (column B)
2. Field of activity (column C)
3. Building area in square meters (column D)
4. Project year (column E)

And optionally:

1. Number of transformers (column I)
2. Transformer power (column H)
3. Number of low-voltage main distribution boards (column N)
4. Number of distribution boards (column Q)
5. Number of inverters (column T)
6. Number of capacitors (column W)
7. Cable tray length (column Y)
8. Number of light fixtures (column AA)
9. Number of switches (column AC)
10. Number of outlets (columns AD and AE)
11. Number of safety devices (column AG)

They would then click on an "Estimate Costs" button and a window would appear containing all the estimated values listed in the other columns. From this window, they could click on a "Similar Projects" button which would list the projects that were closest to the entered project. If possible, an explanation of "how the data was calculated" could be provided (Optional for this mockup).

Model training

A second interface is needed to train the model.

The challenge lies in training the model, which must be trainable by non-IT professionals. The idea is that, when the data (from the 3,500 projects in the Excel file) is submitted, the AI will be able to automatically find correlations and submit them to the user responsible for validating the model.

An example: AI could detect that there was an average inflation coefficient of 3% between 2011 and 2012, find that the number of meters of cable trays is highly correlated to the surface area of the building when we talk about DATA CENTER, detect that the supply / labor ratios are lower for projects in the industrial sector, etc. It could then submit these correlations to the user in charge of the model who would validate them. Training of the model could then continue on these bases.

The software must therefore be able to list these correlations and allow the user responsible for the configuration to take them into account or ignore them. Example: a correlation saying that the power of the transformer station is correlated to the number of fixtures is very likely and will be validated, whereas a correlation saying that it seems that the number of emergency lighting units is linked to the number of capacitors does not make sense and will therefore be rejected by the user in charge of validating the model.

The model training interface could be done in an Excel-like spreadsheet where we would have all the probable correlations identified by the AI and the user would check boxes to indicate what he wants to validate and what he wants to invalidate. The software would then read his responses and train the model accordingly.

Written by : Pierrick Combreau
Date : May, 7 2025

If necessary, the user could add text directives to improve the training of the model, for this the use of an LLM like ChatGPT would probably be necessary.

This model training interface is of course part of the project and will have to work on the mockup.

Deliverables

The successful bidder must provide a complete operational mockup by May 26, 2025, to train the model and use it to predict the costs of new projects. This model will be based on the 3,500 projects provided in the Excel file and does not currently need to be flexible to adapt to new project lists or new areas of activity.

All source code and detailed operating documentation in English, allowing our developers to integrate this model into our software, will be provided upon delivery.

It will be specified if subscriptions are necessary (Copilot? ChatGPT?...) We already have tokens for ChatGPT if needed.

Written by : Pierrick Combreau
Date : May, 7 2025