Figure 2-5. Six Steps to Scope an AI Project – and Shape a Business Case

Understanding the Business Goal

The first step is a rather generic project management task: understanding what the business wants to achieve. It is a clear sign of trust in AI's innovation potential when the business spends half a million or three million euros on implementing an AI solution. Then, they obviously believe that AI helps them run their business better. However, to prevent disappointments, a business case must answer the following questions:

- What is the exact goal of the business? Why do they invest money?

- What are the criteria to decide whether the project delivered everything they expected?

- How does the project relate to strategic and tactical business goals?

- What is the expected timeline?

- Is there already a budget? How much is it?

The answers to these questions help when writing the actual business case. The AI project manager can verify whether the expectations of the management and how the actual high-level project planning after the scoping phase match. Do the project's direction, deliverables, and timeline match the senior management's expectations?

Understand the Insights Category

The second scoping step translates the business problem into an AI problem. Is supervised or unsupervised learning the solution? Which subcategory is needed? Clustering, association, or dimensionality reduction for unsupervised learning or

prediction and classification for supervised learning? Does the training data consist of tables, text, audio, video, or other data types? Figure 2-6 provides a first overview.

Supervised learning algorithms get training data with input and the corresponding correct output. It is always a pair such as a word in English and the proper German translation: <red, rot> <hat, Hut>. One subcategory for supervised learning is **classification**. A classification algorithm puts data items or objects in one of the defined classes. They work similarly to the Sorting Hat in the Harry Potter books and movies. The Sorting Hat determines for each new student in which of the four dormitory houses he or she fits best in and should live in the following years. A typical classification use case is image recognition: Is it a cat in the picture, or is it a dog?

Prediction is the second category of supervised learning insights. How much sugar and cream should we order to produce enough ice cream tomorrow? What will be tomorrow's ice cream sales based on last year's numbers, today's weather, and the weather forecast for tomorrow? Supervised learning is often directly actionable. Thus, they are beneficial for transforming operational business processes in companies and organizations aiming to become data-driven and AI-empowered.
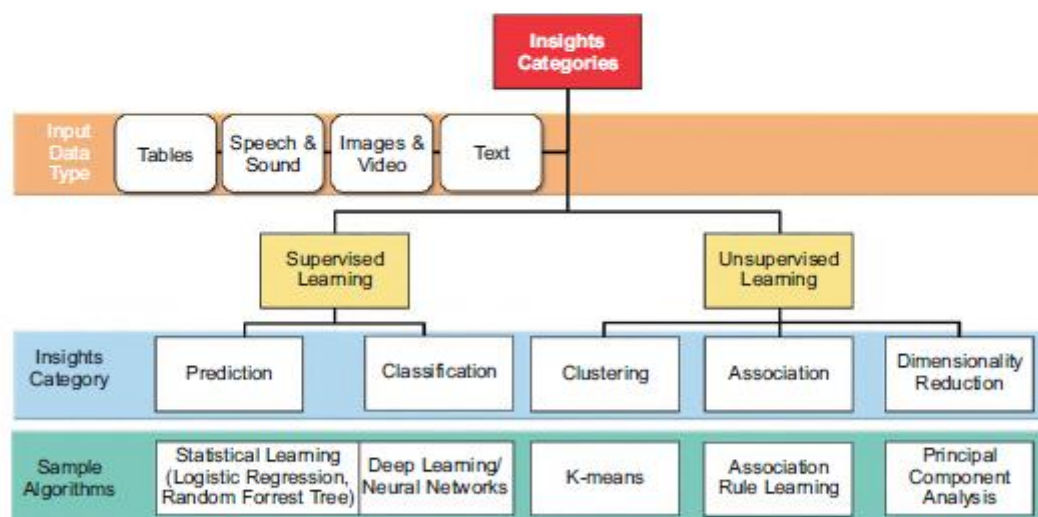

Figure 2-6. *Insights Categories – A Quick Overview*

Unsupervised learning structures the input data. A **clustering** algorithm looks at all data points. It comes up with, for example, three groups of data points that are pretty close together: status-oriented customers, price-sensitive customers, and

nouveau-riche customers. Clustering provides additional insights to the sales and marketing departments and product managers without directly triggering any actions. **Associations** identify relationships between data elements, whereas **dimension reduction** simplifies complex, high dimensional data, for example, to remove noise from images.

In general, unsupervised learning algorithms unveil hidden structures of/in your data, but (often) do not directly tell what to do. While the book uses mainly supervised learning algorithms as examples, many of the methodologies work for unsupervised learning algorithms as well.

When a project has identified the needed insights category, the next step is to understand the data types used for training the model – and where to source them. **Database or CSV/Excel tables** are common formats, though the exact format or file type does not matter much. What matters is that the machine learning algorithms get structured, table-liken input.

Images and videos are other types of input data. They help with camera images during the production process and check whether the produced items have a good quality or video streams from CCTVs to check for persons entering restricted areas. **Speech and sound** (e.g., engine noise) or **text** in the form of documents or emails are additional options.

The exact data types do not impact whether AI can come up with the solution per se. It is more to make a better effort estimation (in short: tables are less work than speech) and elaborate the timeline more precisely. Furthermore, the AI project manager can validate how experienced the data scientists are with the particular area of AI needed for this business question. No project manager wants to realize at halftime that his project misses essential skills, that it is unclear whether suitable AI code libraries exist, or that they must identify a particular external service to build the solution.

Having answers for the questions addressed in the second scoping step brings the project closer to deciding which AI methods and algorithms to use. However, this decision still needs clarity regarding the explainability question.