Six stages of doing a data mining project

Business understanding

Data mining projects begin with business understanding — with companies determining their objectives for a project. Which data does the company wish to study? What are the goals of that study? What problems does the project seek to solve, or what opportunity does it seek to pursue? This stage is essential to determine the right datasets to be analyzed. As a result, data analysts should have a clear understanding of their company's mission, strategy, and objective needs.

Data understanding

With a stated objective, the data mining project moves to the next phase: defining the data. In this step, analysts gather data, describe it (the amount, whether it includes numbers and strings, how it's coded, etc.), and verify its quality. Some key questions for this step: Are there any data gaps? Does the data contain errors? Are fields coded correctly? Is any data duplicated?

It's important to note that not every data point a company stores will fit every project. Gathering the proper data will save time as well as ensure the quality and applicability of insights derived during the project.

Data preparation

Data preparation is often the most time-consuming step of a mining project. In fact, according to IBM, **data preparation can consume** 50-70% of a project's time and effort. Data preparation involves selecting, cleaning, sorting, and formatting the data to be studied. In addition, data from multiple sources will need to be merged or adjusted, and new data may need to be constructed. Once the data has been thoroughly reviewed and prepared, it is ready to be studied.

Modeling

In the modeling stage, data analysts and scientists employ many types of modeling techniques (which we'll explore later) to uncover insights. Perhaps they will run models to find patterns or anomalies. For example, they may run a predictive model to learn whether past data can determine a future outcome. Or, they may run association rule mining (via machine learning models) to discover non-intuitive patterns that provide valuable insights analysts didn't even know were there. It's important to realize that analysts often run multiple models on the same set of data, depending on the project's goals and requirements.

Evaluation

In the evaluation stage, analysts assess whether results answer the business understanding questions properly, meet the project's objectives, or uncover any unexpected patterns. They will also assess whether the correct models were used.

If the initial objective is unmet — or new questions arise — data analysts will return to the modeling phase. In addition the data may need to be adjusted as well. Once the data results answer the business understanding questions, the project reaches its final stage.

Deployment

In the deployment stage, data analysts report their findings and recommend a plan to make those insights actionable. Perhaps the data mining project found that retail customers buy mayonnaise frequently when buying air freshener — a completely non-intuitive insight. With this information, the retailer can craft a marketing plan to take advantage of this insight from a promotional and floor plan perspective.

SAMPLE PROJECTS THAT YOU CAN CHOOSE

The projects are not limited to the following

Fake news detection

Diabetes prediction / Any other disease

House price prediction

Credit card fraud detection

Sentiment analysis

DM project on Customer / financial dataset

DM project on student performance dataset

REFERENCE LINK

https://www.projectpro.io/article/data-mining-projects-ideas-with-source-code/467