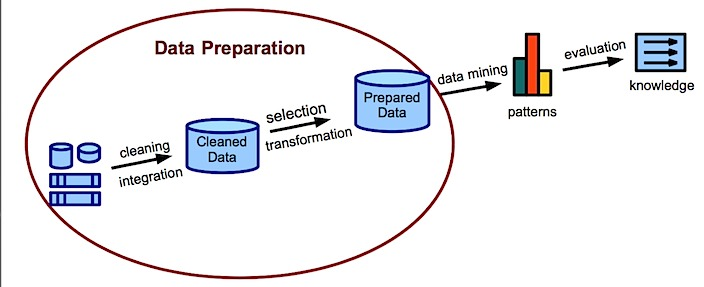
Data Set Preparation

In data analytics and machine learning projects, preparing the data set is an important step. This stage involves tailoring the data to ensure the model learns accurately and effectively.



Combining Training and Test Data Sets

We combined the training and test data sets to be used in the project. By this merging process we ensured that they have the same columns and contain the same type of information.

In this way, a main data set that will have the same data structure for model training and prediction processes is obtained.

Editing Date Data

Historical data is critical for time series analysis. Therefore, it is important to convert date data into appropriate format and extract relevant features.

With these steps, we made date data actionable. For example, we had access to information such as which month and day of the week each sale took place.

Removing Unnecessary Columns

Columns in the data set that are not required for modeling or contain redundant information are removed. This step makes the model less complex and more effective.

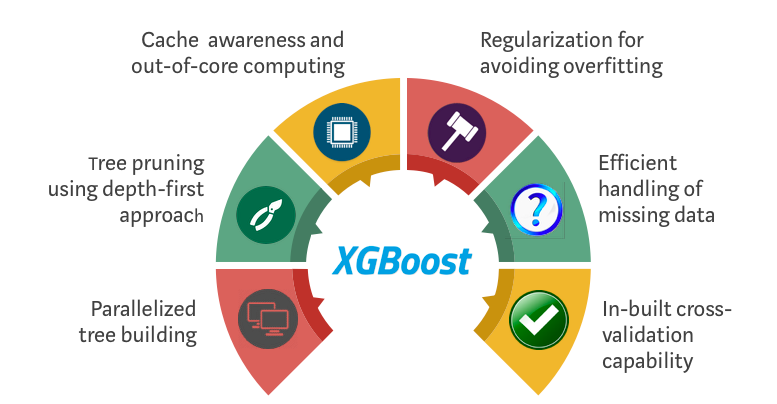
In this way, we obtained a data set consisting only of columns that will be used in the modeling process.

All of these steps ensure that the machine learning model is trained correctly, making the data more organized, understandable, and ready for modeling.

Model Training

We chose to use XGBoost in our project. As for its features, we would like to briefly explain it.

XGBoost (eXtreme Gradient Boosting) is a powerful machine learning algorithm that is used especially on structured data sets and achieves successful results. It is a tree-based model based on the Gradient Boosting method.



Creating an XGBoost Regression Model

In this step, we create the regression model available in the XGBoost library.

By changing the max\_depth parameter to different values, we control the complexity of the model.

Performance Evaluation

For each model, we made predictions on the training and testing datasets.

We calculated SMAPE (Symmetric Mean Absolute Percentage Error) and MSE (Mean Squared Error) values, which are the success criteria of the predictions.

Selection of the Best Model

As a result of the performance evaluation, we choose the best model. For example, here we chose the model at index 2.

We retrain the best model on the full dataset.

These steps are important to get the best performance by trying different model parameters. The best model chosen will be one that has high predictive ability and can generalize the data accurately. This detailed evaluation during the model training phase plays a critical role in increasing project success.