

```
<scenario>
<AnalysisScenario>
  <Step name="features selection">
    <Task id="1" name="electrical current features selection
with hierarchical clustering" />
  </Step>
</AnalysisScenario>
</scenario>
```

Consider the definitions (between <definitions>  
</definitions>) of step and task as well as of inclusion,  
extension and specialisation between pairs of analysis scenarios.

<definitions>

An analysis scenario is composed of steps and each step is  
accomplished through a series of tasks. You can think of a step as  
a macro-group of affine tasks.

Specialisation relationship between scenarios: indicates that a  
scenario adds several steps/tasks with respect to another  
scenario.

Extension relationship between scenarios: to emphasise that a  
scenario presents a list of steps/tasks that is a variation of the  
steps/tasks of another scenario.

Inclusion relationship between scenarios: indicates that a  
scenario is included in other scenarios by specifying that the  
steps/tasks within one scenario are inclusive of the steps/tasks  
in other ones.

</definitions>

Based on the definitions between <definitions> </definitions>, and  
the analysis scenarios provided in the uploaded files, try to  
complete the scenario taking into account the scenario in the  
given context, the definitions of inclusion, extension and  
specialisation and the textual description of the whole scenario.

If you are not able to answer say "I don't know".

Analysis scenario description:

I have to read a database table, containing 10 columns: 1 column for  
the record ID, 1 column for the variable short name, 1 column for a  
timestamp value, 6 columns containing electrical current values and a  
column containing torque values (in Newton meter) collected from a  
smart machine engine.

The columns containing the record ID and the variable short name are  
not required for the next analysis steps.

I want to use the 6 columns containing electrical current values for  
building a LASSO regression model and the torque has to be the  
prediction target for the model.

Afterwards, I want to compare the predicted torque values with the real values.

Based on the definitions and the provided analysis scenarios, here is the completed scenario incorporating the information given:

```
<scenario>
<AnalysisScenario id="AS4" extends="AS1.xml">
  <Step id="1" name="data collection" supervisedBy="Client
company">
    <Task id="1" name="read database table with electrical
current and torque values" />
    <Task id="2" name="extract relevant columns (electrical
currents and torque)" />
  </Step>
  <Step id="2" name="data preprocessing" supervisedBy="Client
company">
    <Task id="1" name="remove unnecessary columns (record ID
and variable short name)" />
    <Task id="2" name="handle missing values if any" />
  </Step>
  <Step id="3" name="features selection" supervisedBy="OEM">
    <Task id="1" name="electrical current features selection
with hierarchical clustering" />
  </Step>
  <Step id="4" name="torque prediction" supervisedBy="OEM">
    <Task id="1" name="create dataset with selected features
(electrical currents) and torque" />
    <Task id="2" name="train-test dataset split" />
    <Task id="3" name="LASSO regression model building" />
    <Task id="4" name="calculation of predicted torque values"
/>
  </Step>
  <Step id="5" name="comparison of predicted vs real torque
values" supervisedBy="OEM">
    <Task id="1" name="calculation of pearson correlation" />
    <Task id="2" name="calculation of mean root squared error"
/>
  </Step>
</AnalysisScenario>
</scenario>
```

Explanation:

- **Data Collection:** This step involves reading the database table and extracting the necessary columns. Tasks related to extracting relevant columns and reading the table are added.

- Data Preprocessing: This step includes removing unnecessary columns and handling missing values, preparing the data for analysis.
- Features Selection: Retains the task for selecting electrical current features with hierarchical clustering as provided.
- Torque Prediction: Adds tasks for creating the dataset, splitting it into train-test sets, building the LASSO regression model, and calculating predicted torque values.
- Comparison of Predicted vs Real Torque Values: Involves calculating the Pearson correlation and mean root squared error to compare predicted and real torque values.

This scenario extends the existing AS1.xml scenario by adding specific tasks related to reading the database, preprocessing the data, and using it for torque prediction and comparison.