Imagine a biology student who needs to analyze and evaluate data measured in a lab. For her, as a domain expert in her field, the business and the data understanding phases of the data mining process are not a problem. The main challenge for her is to pre-process and analyze the data and gain useful knowledge from it.

Our work is aimed at helping people to analyze their data in a simple and user-friendly way with no previous knowledge of data analysis nor data-mining.

**How to guide a non-technical user through the whole data-mining process?**

The solution should be:

- user-friendly
- easy to use
- fast (no long waitings for results)
- accurate

**Results of our hyper-parameter search approach**

**ERROR RATE**

<table>
<thead>
<tr>
<th>13,000</th>
<th>10,400</th>
<th>7,800</th>
<th>5,200</th>
<th>2,600</th>
<th>0</th>
</tr>
</thead>
</table>
| 5.27%  | 3.07%  | 6.73% | 9.29% | 1.30% | 0%

† Sum of positions of recommended models for different recommendation techniques on 52 datasets (183 484 models total).

**TIME COMPLEXITY**

<table>
<thead>
<tr>
<th>1,500</th>
<th>350</th>
<th>100</th>
<th>60</th>
<th>40</th>
<th>20</th>
<th>0</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.27%</td>
<td>3.07%</td>
<td>6.73%</td>
<td>9.29%</td>
<td>1.30%</td>
<td>0%</td>
<td></td>
</tr>
</tbody>
</table>

† Number of trained models for different model selection algorithms.

Our proposal:

- Automatic conversational system pre-processes the input data and generates questions for user to determine the further steps.
- Hundreds of differently preprocessed data files using computational cluster for fast and reliable results.
- Custom Meta-learning algorithm with Landmarking features speeds up the combined model and hyperparameter selection.

**More Results**

We focused on classification problems in our current working prototype.

DEMO: [http://s.ics.upjs.sk/~sbocko/winston](http://s.ics.upjs.sk/~sbocko/winston)

We will support regression and pattern mining techniques soon. Planned public release of this software is on 1st of August 2015.

Using the Meta-learning we were able to speed up the model recommendation time. For this purpose we combined our Landmarking approach with statistical characteristic approach presented by R. Neruda et al.