Data Analysis System Improvements

Levent Güner
01.11.2019
Improvements

• Loading page
• Allowance to multiple users (for server)
• Visualization improvements
• Gateway to missing value prediction
• Missing value prediction’s back-end is almost done, next step is to run with front-end
• Demo:
  • http://leventguner.net/parton/partonrep_6432.html
  • http://leventguner.net/parton/partonrep_8921.html
### Completed Date

**Categorical:** 51.14%

<table>
<thead>
<tr>
<th>Tab</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overview</td>
<td></td>
</tr>
<tr>
<td>Missing Values</td>
<td>Missing values: 51.14%</td>
</tr>
<tr>
<td></td>
<td>Missing values (n): 101709</td>
</tr>
</tbody>
</table>

**Predict/Impute missing values according to...**

- all values
- closest 3 columns
- most common values

[Do Magic!]

### First Construction Document Date

**Categorical:** 7.51%

<table>
<thead>
<tr>
<th>Tab</th>
<th>Data</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overview</td>
<td></td>
</tr>
<tr>
<td>Missing Values</td>
<td>Missing values: 7.51%</td>
</tr>
<tr>
<td></td>
<td>Missing values (n): 14946</td>
</tr>
</tbody>
</table>

**Predict/Impute missing values according to...**

- all values
- closest 3 columns
- most common values

[Do Magic!]
DataWig - Imputation for Tables

• DataWig learns Machine Learning models to impute missing values in tables
• DataWig expects us to provide the column we would like to impute values for and some columns that contain values that we deem useful for imputation
• Possible to use with GPU
Next steps

• Using DataWig ([https://github.com/awslabs/datawig](https://github.com/awslabs/datawig)) to predict categorical / non-numeric missing values
• Adding more tables and features
• Completing benchmarking for different methods and visualizing them