I. Problem Statement

Create (i) Benchmarking Suite and (ii) Visual plugins for ‘empowering’ ECL-ML to make development faster and more reliable.

II. Why do we need these tools?

a) ECL is easy, but not so easy!
   • ECL has a steep learning curve
   • ECL-ML requires end-user to understand the details of the library, which needs extensive documentation

b) Current Workflow can be tedious!
   • Benchmarking Suite
     • Compares performance of ECL-ML library with scikit-learn
     • Supports Software Regression Testing of ML algorithms
   • Testing Suite found regression error in Logistic Regression
   • Time required to run the suite is ~4 hours

b) ECL-ML Plugins
   • Integrated with Data Science Portal (DSP)
   • Access to distributed ML Algorithms
     • Classification
     • Regression
     • Clustering
   • Workflow based development
   • Supports
     • Quick data visualization support

III. Contributions

a) Benchmarking Suite
   • Compares performance of ECL-ML library with scikit-learn
   • Supports Software Regression Testing of ML algorithms
   • Extensible framework supports test cases for new ML algorithms

<table>
<thead>
<tr>
<th>Tasks</th>
<th>Algorithms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Classification</td>
<td>Random Forest</td>
</tr>
<tr>
<td></td>
<td>Decision Trees</td>
</tr>
<tr>
<td></td>
<td>Logistic Regression</td>
</tr>
<tr>
<td>Regression</td>
<td>Linear Regression</td>
</tr>
<tr>
<td>Clustering</td>
<td>KMeans</td>
</tr>
<tr>
<td>Time Series</td>
<td>ARIMA</td>
</tr>
</tbody>
</table>

IV. Future Work

• Integrate the ML-Regression Suite to Platform Regression Suite
• Add more algorithms and make it available for production