Breakthrough Automatic
Machine Learning Technology

Edammo AutoML Technology Benchmark Analysis

Brandon Warner
Edward Ratner

Edammo Inc

1. Overview of Novel Approach

In the context of machine learning, performance is measured by a combination of scoring metrics (accuracy, precision, etc.) and training time. Many contemporary machine learning technologies, such as Deep Learning, require high volumes of data and long training times. Edammo AutoML has approached machine learning classification, regression, and time-series prediction in a fundamentally different way. This allows Edammo's AutoML technology to produce fast, accurate models on small data sets. This novel architecture enables businesses of any size to leverage their dynamic, ever-changing data in real time.

Classical and Deep Learning algorithms also require extensive hyperparameter optimization. Edammo AutoML's revolutionary architecture automatically produces highly accurate models without the need to optimize hyperparameters, allowing customers to instead focus their efforts on leveraging these highly accurate models to their advantage.

Another important variable to consider when choosing the best machine learning technology for classification problems is the performance across classes. While a model with relatively high accuracy may seem adequate at first, poor performance on one or more of the classes will result in a significant decrease in utility. The metric that best measures performance across classes is the Jaccard Index:

$$J(A, B) = \frac{|A \cap B|}{|A \cup B|}$$

where *J* is the Jaccard Distance between sets *A* and *B*

As shown in the tables below, Edammo AutoML's novel algorithmic approach not only outperforms industry leaders in terms of total accuracy, but also produces higher minimum Jaccard Indices, illustrating greater generalization. We currently report the results for classification. In the coming months we will add results for the general regression problem as well as time series.

2. Experiments

The University of California at Irvine's machine learning repository contains thousands of open-source datasets used for classification, regression, time-series, and clustering experiments. To best benchmark Edammo AutoML's technology to state-of-the-art industry leaders, performance comparisons were made on classification models side by side with Google's AutoML service. Metrics measured include accuracy, training time, and Jaccard Indices for each class in the dataset. The first experiment utilized the Human Activity Recognition Using Smartphones dataset (https://archive.ics.uci.edu/ml/datasets/human+activity+recognition+using+smartphones), which was separated into predefined training testing splits (80% training, 20% testing) and consists of 561 attributes and 10,299 observations. The second dataset, Parkinson's Disease Classification (https://archive.ics.uci.edu/ml/datasets/Parkinson%27s+Disease+Classification), was not split beforehand. Therefore, 5-fold validation was used (five splits of 80% training, 20% testing). The original Parkinons's dataset is 754 variables by 756 observations, but since Google AutoML requires a minimum of 1000 observations to train a model, the dataset was doubled and 0.1% noise was added. The third experiment used the QSAR Oral Toxicity (https://archive.ics.uci.edu/ml/datasets/QSAR+oral+toxicity) dataset, also using 5-fold validation. However, Google AutoML failed to train on two of the five folds so only the successfully predicted folds were included in the comparison. The fourth dataset, CNAE9 (https://archive.ics.uci.edu/ml/datasets/cnae-9), failed to train using Google AutoML so the dataset was doubled and 0.1% of noise was added to the training sets. Only one of the five folds successfully trained using Google AutoML so only that fold was included in the comparison.

3. Results

Human Activity Recognition

Figure 1 shows the breakdown of the Human Activity Recognition dataset's attributes and the performance for each learning method. Edammo's AutoML technology has the ability to be optimized for accuracy or speed. In the first experiment using the Human Activity dataset, the Edammo AutoML fast method outperformed Google AutoML by 1.66% accuracy, whilst training the model over 72 times faster. In this case, the fast method was just as accurate as the accurate method so only the fast method was included in the comparison. Figure 2 shows the breakdown of the various Jaccard Indices for the six classes in the Human Activity Recognition dataset. For each of the classes, Edammo AutoML outperformed Google AutoML in terms of Jaccard Indices, showing greater generalization capabilities.

Table 1. Results Table 1 for Human Activity Dataset

Data	Method	Split type	Training Instances	Attributes	Training time (seconds)	Accuracy
Human Activity	Edammo AutoML fast and accurate method	Predefined 1-fold	8239	561	203	**0.9626**
Human Activity	Google AutoML	Predefined 1-fold	8239	561	14700	**0.946**

Table 2. Jaccard Indices Table for Human Activity Dataset

Data	Method	Jaccard Class 1	Jaccard Class 2	Jaccard Class 3	Jaccard Class 4	Jaccard Class 5	Jaccard Class 6
Human Activity	Edammo	0.944	0.9476	0.9718	0.8613	0.875	0.9816
Human Activity	Google AutoML	0.9301	0.9109	0.9222	0.8585	0.8727	0.9795

Parkinson’s Disease Classification

In the second experiment using the Parkinson’s dataset, Edammo AutoML accurate method outperformed Google AutoML by 1.45% accuracy, whilst training almost 30 times faster. The Edammo AutoML fast method slightly outperformed Google AutoML by 0.2%, but at a significantly faster training speed (291 times faster than Google AutoML). Figure 4 shows the breakdown of the Jaccard Indices for the two classes in the dataset. For each of the classes and methods, Edammo AutoML outperformed Google AutoML in terms of Jaccard Indices, further illustrating greater generalization capabilities.

Table 3. Results Table for Parkinson’s Disease Classification Dataset

Data	**Method**	**Split type**	**Training Instances**	**Attributes**	**Training time (seconds)**	**Accuracy**
Parkinson's	Edammo AutoML accurate method	5-fold	1202	752	2050	**0.9233**
Parkinson's	Edammo AutoML fast method	5-fold	1202	752	205	0.9088
Parkinson's	Google AutoML	5-fold	1202	752	59820	**0.9046**

Table 4. Jaccard Indices Table for Parkinson’s Disease Classification Dataset

Data	**Method**	**Jaccard Index for Class 1**	**Jaccard Index for Class 2**
Parkinson's	Edammo AutoML Accurate Method	0.85416	0.86158
Parkinson's	Edammo AutoML Fast Method	0.82848	0.8371
Parkinson's	Google AutoML	0.8217	0.8302

QSAR Oral Toxicity

In the third experiment using the QSAR Oral Toxicity dataset, the Edammo AutoML accurate method performed 2.17% better than Google AutoML, whilst training 19 times faster. The Edammo AutoML fast method outperformed Google AutoML by 1.93%, whilst training over 194 times faster. Figure 5 shows the breakdown of the Jaccard Indices for each of the classes in the QSAR Oral Toxicity model. For each of the classes and methods, Edammo AutoML outperformed Google AutoML in terms of Jaccard Indices. Most notably, in this unbalanced dataset, the Jaccard index for the underrepresented class was more than 2x higher for the Edammo AutoML model than the Google AutoML model.

Table 5. Results Table for QSAR Oral Toxicity Dataset

Data	Method	Split type	Instances	Attributes	Training time (seconds)	Accuracy
QSAR Oral	Edammo AutoML accurate method	3-fold	4316	700 (Google failed on 999, 900, 800)	1920	**0.9394**
QSAR Oral	Edammo AutoML fast method	3-fold	4316	700 (Google failed on 999, 900, 800)	192	0.937
QSAR Oral	Google AutoML	3-fold	4316	700 (Google failed on 999, 900, 800)	37260	**0.9177**

Table 6. Jaccard Indices Table for QSAR Oral Toxicity Dataset – underrepresented class bolded

Data	Method	Jaccard Index for Class 1	Jaccard Index for Class 2
QSAR Oral	Edammo AutoML Accurate Method	**0.3853**	0.937
QSAR Oral	Edammo AutoML Fast Method	0.3522	0.934
QSAR Oral	Google AutoML	**0.1868**	0.9161

CNAE9 Business Classification

In the final experiment using the CNAE9 dataset, Edammo AutoML performed 1.4% better than Google AutoML, whilst training 165 times faster. Figure 8 shows the breakdown of the Jaccard Indices for each of the classes in the CNAE9 Business Classification model. Edammo AutoML's minimum Jaccard Index was higher than Google AutoML by 0.1, a 20% improvement, showing greater generalization capabilities.

Table 7. Jaccard Indices Table for CNAE9 Business Classification Dataset

Data	**Method**	**Split type**	**Training Instances**	**Attributes**	**Training time (seconds)**	**Accuracy**
CNAE9	Edammo AutoML	1-fold	1728	856	83	**0.7721**
CNAE9	Google AutoML	1-fold	1728	856	13740	**0.7581**

Table 8. Jaccard Indices Table for CNAE9 Business Classification Dataset – minimums bolded

Data	**Method**	**Jaccard Class 1**	**Jaccard Class 2**	**Jaccard Class 3**	**Jaccard Class 4**	**Jaccard Class 5**	**Jaccard Class 6**	**Jaccard Class 7**	**Jaccard Class 8**	**Jaccard Class 9**
CNAE9	Edammo AutoML	0.7692	0.5454	0.5937	0.5714	0.7619	**0.5**	0.6667	0.7037	0.6285
CNAE9	Google AutoML	0.7916	0.7407	0.4038	0.5625	1.0000	0.4706	0.8800	0.7037	**0.4000**

Visualizations of Results

Relative error rates are visualized below (figure 9) in order to best illustrate the varying classification performance between Google AutoML and Edammo AutoML. Due to the vast difference in training times between the two services, a dual-axis bar chart featuring a 5x scale increase on the Edammo AutoML axis was employed to improve clarity.

Fig. 1. Visualization showing relative error rates for all datasets

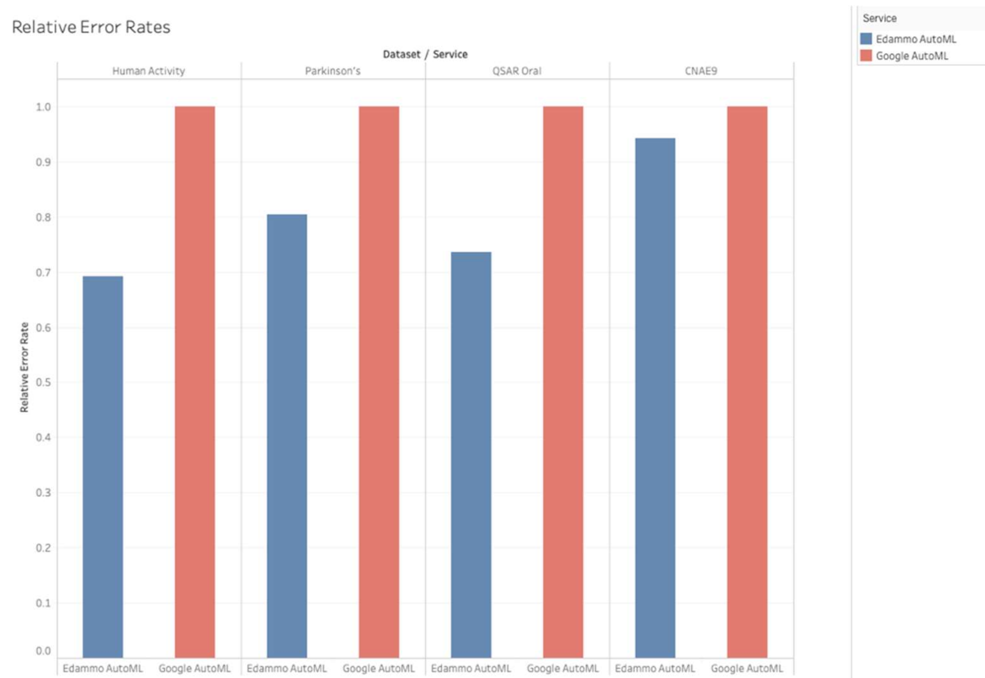

Fig. 2. Training times for all datasets

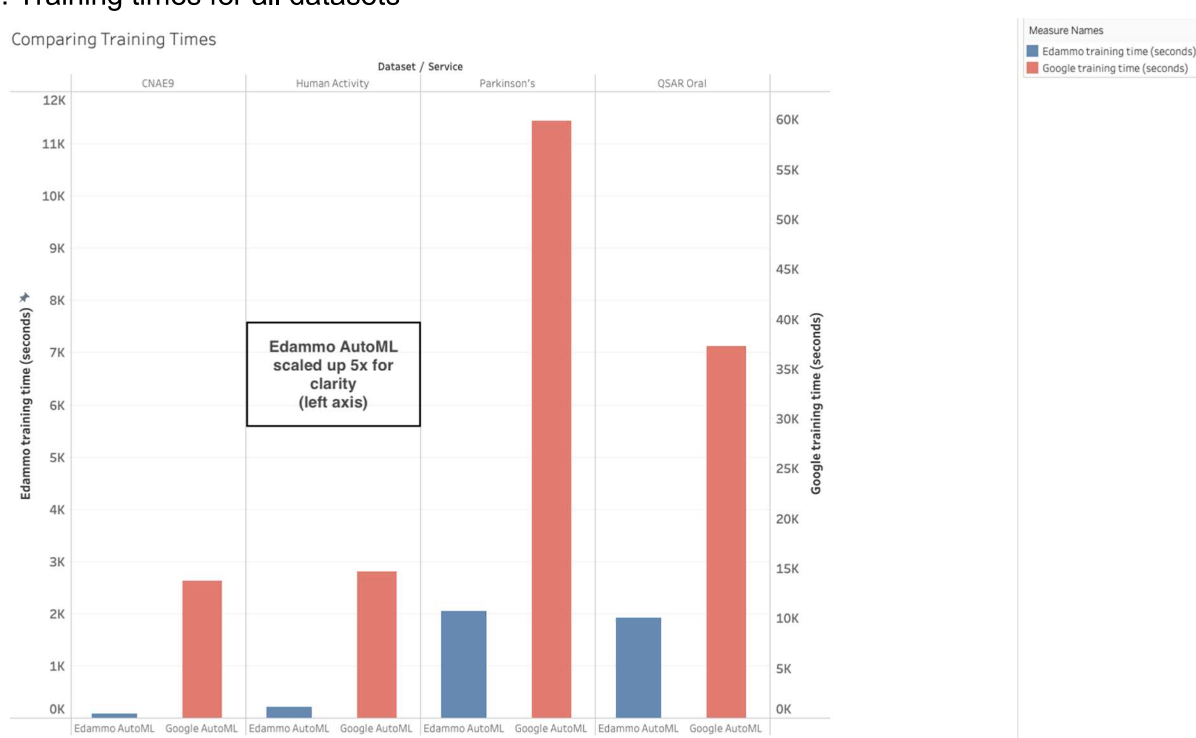

Fig. 3. Visualization showing accuracy and Jaccard Indices for Human Activity Dataset

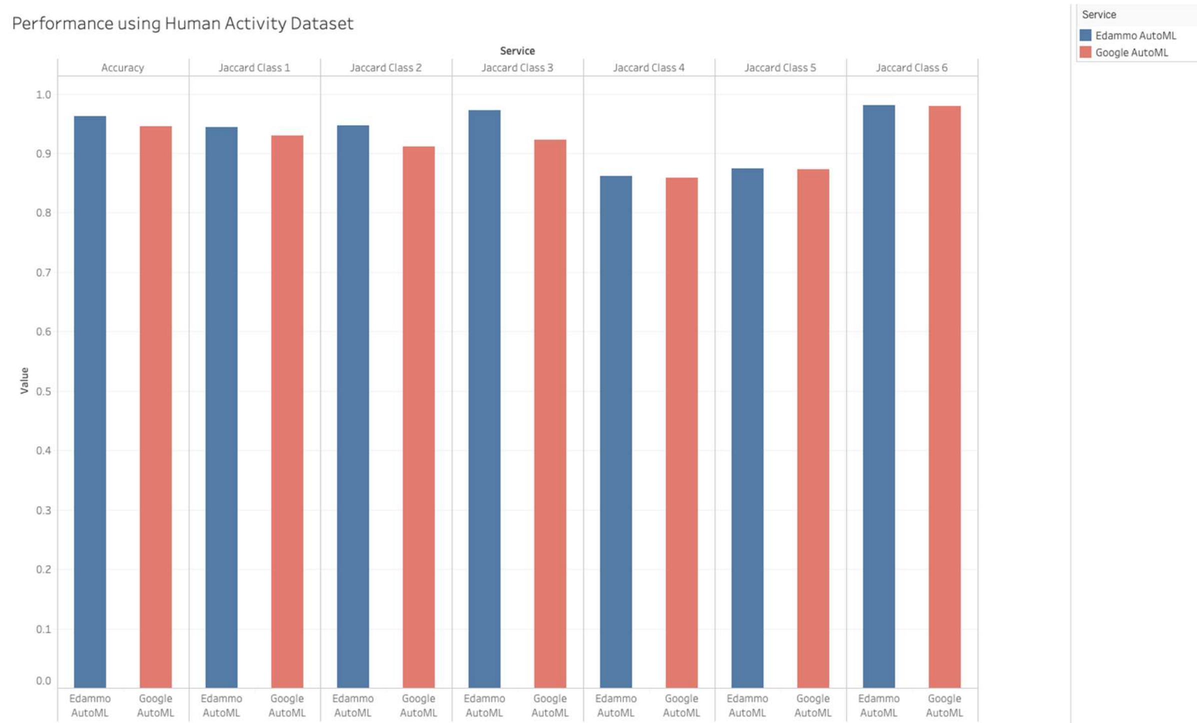

Fig. 4. Accuracy and Jaccard Indices visualization using Parkinson's Disease Classification Dataset

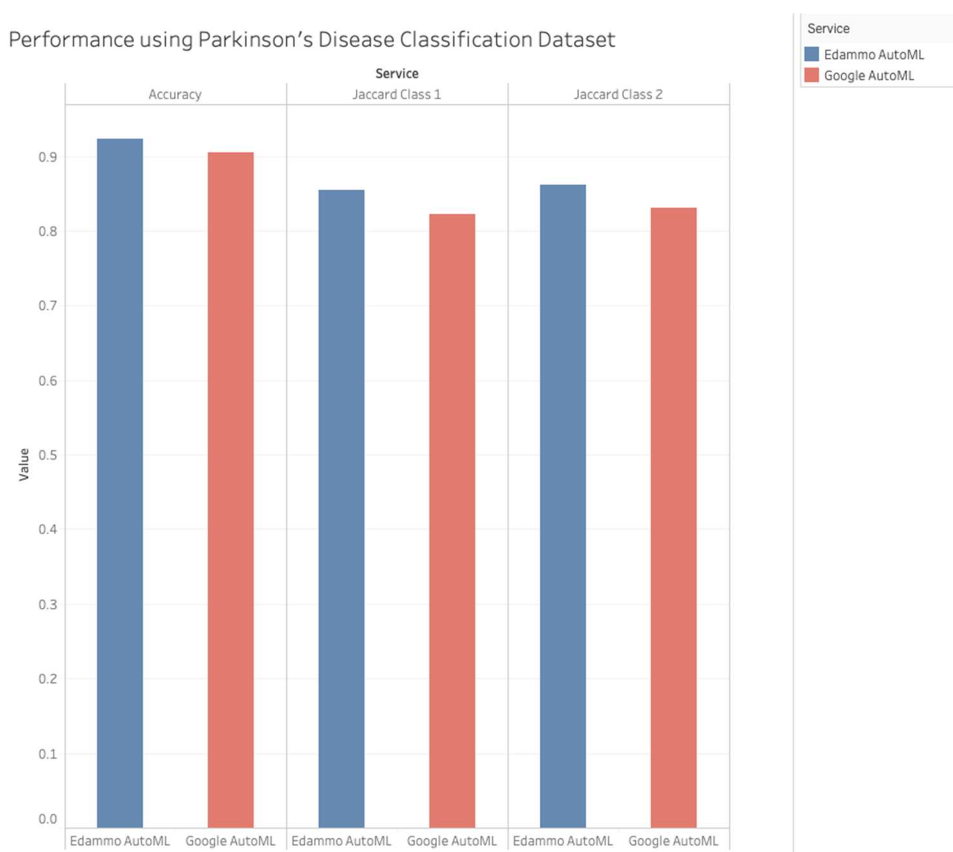

Fig. 5. Visualization showing accuracy and Jaccard indices for QSAR Oral Toxicity dataset

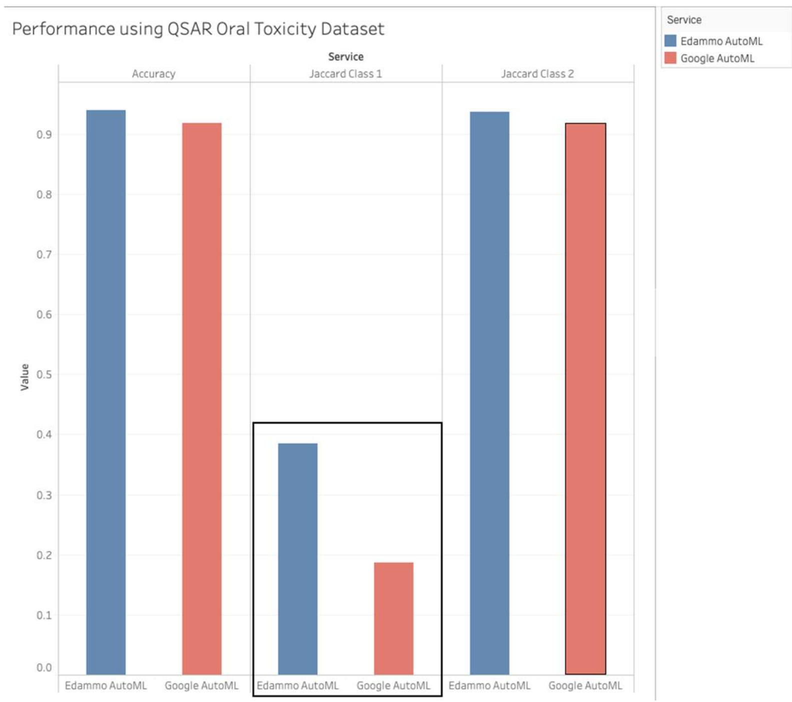

Fig. 6. Visualization showing accuracy and Jaccard indices for CNAE9 dataset