



White Paper

# Best of Both Worlds with Continuous Machine Learning: Do More with Less Computational Resources

**Tazi.ai**

353 Sacramento St STE 1800

San Francisco, CA 94111

+1(408)290-5491

info@tazi.ai



## SUMMARY

Continuous machine learning technology allows, if desired, updating the machine learning system at every instance. In a continuous frame of operation, we accept that change is inevitable. Models updated automatically at every instance are always ready, make less mistakes and allow more business benefits. These models can be used directly in deployment, or for continuous monitoring. When the deployed model needs to be updated, the continuous monitoring model can replace it without training and through much faster validation procedures.

On stock market data, we compare batch and continuous machine learning and demonstrate efficiency (the computational resource benefits-CPU time) and accuracy of continuous learning compared to batch machine learning decision trees and XGBoost. For 3 months of stock data, the continuous machine learning system updated every 1 minute, has similar accuracy to XGBoost updated every 10 minutes and is 127 times faster. The continuous machine learning system updated every 1 minute, has better accuracy than decision trees updated every 10 minutes and is 23 times faster. Considering the fact that businesses need more machine learning solutions for different tasks and the world keeps changing in unpredicted ways and always, the cloud cost savings using continuous learning will be quite significant. Continuous learning will also allow democratization of AI by letting small players benefit from AI using less infrastructure costs.

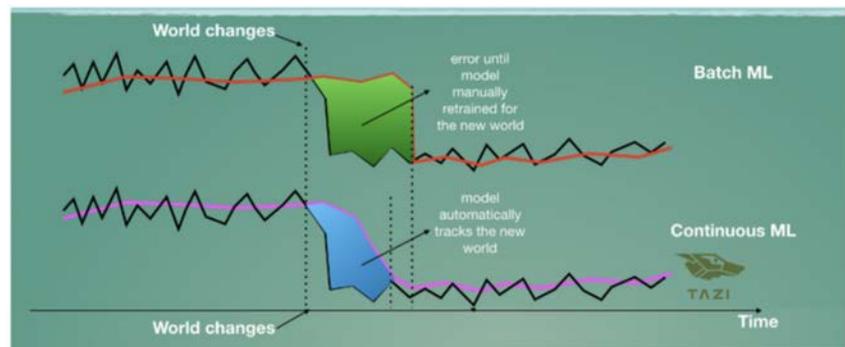


Figure 1: Automated Churn Reduction Approach



## INTRODUCTION

As humans, we keep learning during our lifetime, from everything we are exposed to and at the time we are exposed to. But when traditional batch machine learning (batch ML) methods are used, a labeled dataset is collected, model(s) are trained on that dataset. The available data is usually divided into a training and test set. A model is trained on the training set, i.e. the parameters of the model are optimized to fit the training data. Then the trained model is tested on the test set. If the test performance is good, then it is assumed that the performance of the deployed model would also be good. (see i.i.d., independent and identically distributed, assumption). The deployed model predicts the target for every input (instance) it is presented with.

However, as customers' behavior, economy, products, competitors change so does the data. In batch ML since the deployed model is not updated, it goes stale when data changes. Business (users of ML) realize that the model has more errors than usual, warns the data science team, who try to understand what is happening. Usually there is a wait until enough data is collected. A new model is trained and tested from scratch and deployed. Between the beginning of change and deployment of the new model, errors increase causing revenue losses (See Figure 1). Training and deployment of new models cause pressure on the IT and data science teams, especially if a large number of machine learning models all need to be updated at the same time. In some cases, the model rebuilding is a periodic task to keep the models up-to-date. However, re-training of the models can be, computationally and person/calendar time-wise, quite an expensive task since a large data set has to be processed each time.

With continuous machine learning technology, models are updated as soon as data become available. Models are not trained from scratch, but updated incrementally. The update happens automatically and for each instance. Compared to batch learning that starts from scratch to re-train a model, the time to only update (i.e. not re-train) the continuous machine learning models is a lot less than batch learning.



## Data and Task

We compare the time it takes to train and test machine learning models used in a continuous learning (update the model at every instance) or batch learning (at every k instances, e.g. 10 or 60), train the model from scratch using all the available data so far. For continuous learning, TAZI AutoML v2.3.1. with continuous learning decision tree and a continuously updated explanation model (another decision tree) have been used. For batch learning, Python scikit-learn decision tree and XGBoost classifiers have been used.

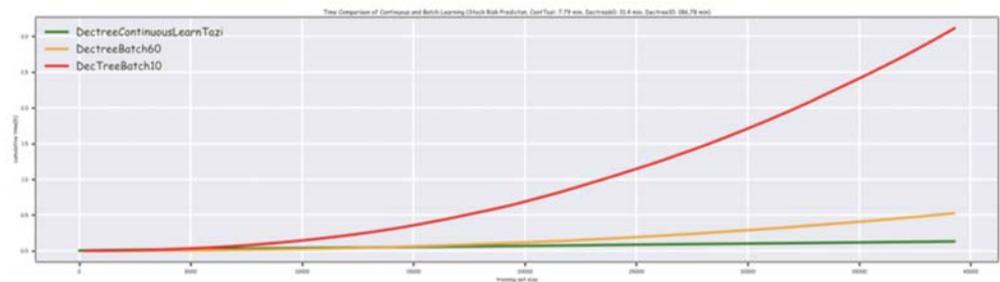
**Task:** Predict if the stock price will drop significantly within the next 1 min.

**DataSet:** Borsalstanbul Garanti Bank Stock, 3 months data, 40000 instances (10.10.2019-06.02.2020), 46 features.

## Time comparison of batch machine learning vs continuous machine learning

XGBoost vs Continuous Learning:

The total CPU time spent for both training and prediction tasks for the whole dataset is:



### Tazi.ai

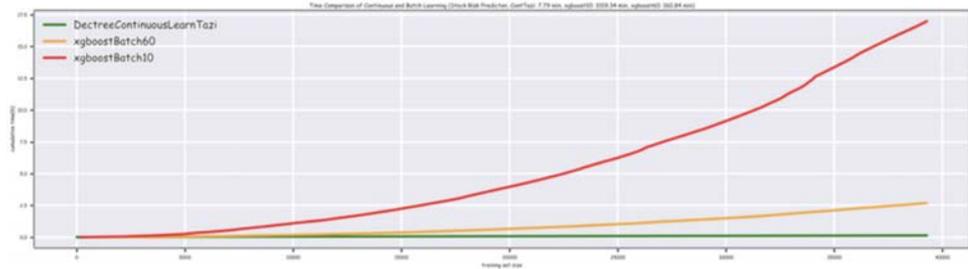
353 Sacramento St STE 1800  
San Francisco, CA 94111  
+1(408)290-5491  
info@tazi.ai



### Decision Tree vs Continuous Learning:

The total spent for both training and prediction tasks for the whole dataset is:

Continuous Learning every instance (DectreeContinuousLearnTazi) :	8 mins
Batch Learning every 60 instances (DectreeBatch60) :	31 mins
Batch Learning every 10 instances (DectreeBatch10) :	3hours 7 mins



### Accuracy comparison of batch machine learning vs continuous machine learning

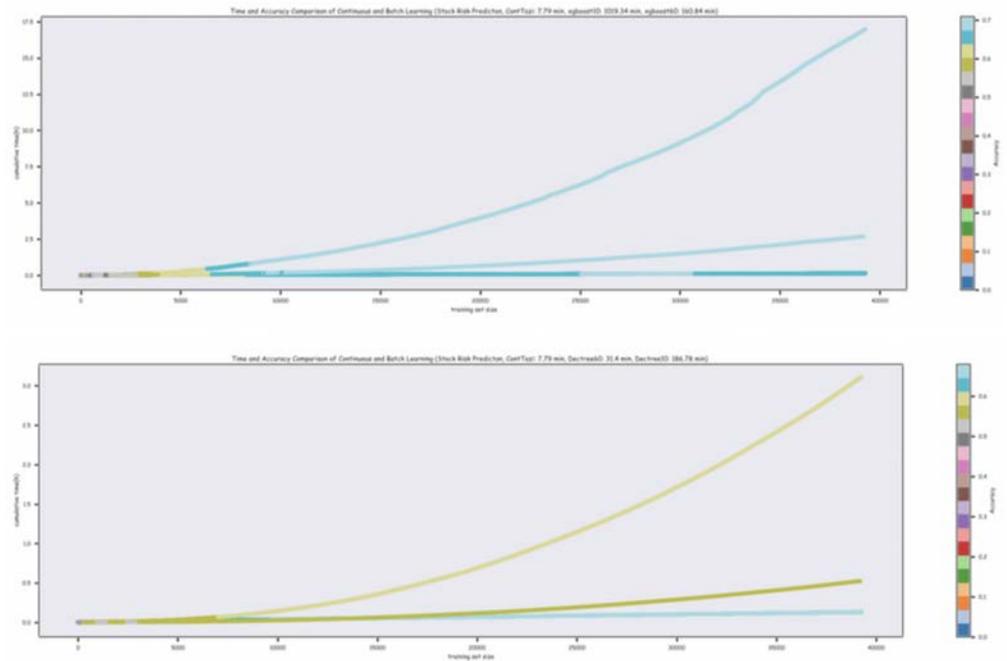
XGBoost has been frequently used in the machine learning community due to its accuracy. Here we compare the accuracy of continuous machine learning to decision trees and XGBoost. The top and bottom figures show the accuracies for the experiments we reported in Section 3. Colors indicate (see legend) the accuracy level achieved by each classifier for the instances in that time period.

XGBoost, indeed, achieves better accuracy than the decision tree. But it is slower than the decision tree. (Compare top and bottom figures, the curve at the top in each figure.)

On the other hand, continuous machine learning is more accurate than decision tree (bottom figure) and it achieves as much accuracy as XGBoost (top figure).

#### Tazi.ai

353 Sacramento St STE 1800  
San Francisco, CA 94111  
+1(408)290-5491  
info@tazi.ai



### Accuracy comparison of batch machine learning vs continuous machine learning

DETAILS OF THE EXPERIMENTS:

Label: Label is calculated based on whether the stock price is higher than the future exponential moving average.

An exponential moving average (EMA) is a type of moving average (MA) that places a greater weight and significance on the most recent data points.

For a series EMA may be calculated as follows:

$$S_t = \begin{cases} Y_1, & t = 1 \\ \alpha \cdot Y_t + (1 - \alpha) \cdot S_{t-1}, & t > 1 \end{cases}$$



Where:

- The coefficient  $\alpha$  represents the degree of weighting decrease, a constant smoothing factor between 0 and 1. A higher  $\alpha$  discounts older observations faster.
- $Y_t$  is the value at a time period  $t$ .
- $S_t$  is the value of the EMA at any time period  $t$ .

If the closing price of the stock at time  $t$  is higher than the EMA of the next 3 minutes, the label is defined as HighRisk.

#### **Experiment Setup:**

The experiment was run on a MacBook Pro laptop computer with the following specification:

Processor Name: Intel Core i5

Processor Speed: 2.3 GHz

Number of Processors: 1

Total Number of Cores: 2

Memory: 16 GB

#### **Decision Tree & XGBoost Algorithm:**

Python version: Python 3.7.0

Scikit-learn version 0.20.3

#### **Tazi.ai**

353 Sacramento St STE 1800

San Francisco, CA 94111

+1(408)290-5491

info@tazi.ai



## ABOUT TAZI

Artificial intelligence (AI) is a source of both huge excitement and apprehension, transforming enterprise operations today. It is more intelligent as it unlocks new sources of value creation and becomes a critical driver of competitive advantage by helping companies achieve new levels of performance at greater scale, growth, and speed than ever before, making it the biggest commercial opportunity in today's fast changing economy.

TAZI is a leading global Automated Machine Learning product/solutions provider with offices in San Francisco. TAZI is a Gartner Cool Vendor in Core AI Technologies (May 2019) and is considered as "[The Next Generation of Automated Machine Learning](#)" by Data Science Central.

## WHO WE ARE

Founded in 2015, TAZI has a single mission which is to help businesses to directly benefit from Automated Machine Learning by using TAZI as a superpower, shaping the future of their organizations while realizing direct benefits like cost reduction, increasing efficiency, enhanced (dynamic) business insight, new business (uncovered), and business automation.

## WHAT WE OFFER

Through its understandable continuous machine learning from data and humans, TAZI is supporting companies in banking, insurance, retail, and telco industries in making smarter, more intelligent business decisions.



TAZI solutions are based on a most compelling architecture that combines the experiences of 23 patents granted in AI and real-time systems, proven at different global implementations.

Some unique differentiators of TAZI products are:

- Business users can automatically configure custom ML models based on their KPI and the available data. TAZI's Profiler accelerates this process through data understanding and automated cleaning, feature transformation, engineering, and selection capabilities.
- TAZI models learn continuously, and are suitable for today's dynamic, realtime data environments.
- TAZI models are GDPR compliant (no black-box models). They provide an
- explanation in the business domain's terminology for every result they produce.
- TAZI supports multiple (heterogeneous) data sources, i.e.,...: external, batch, streaming, and others.
- TAZI can learn both from human domain experts and from data, which speeds up accuracy improvement.
- TAZI's hyper parameter optimization feature reduces human time spent for model configuration. TAZI products contain algorithms that are developed and coded to be lean, efficient, and scalable.

**Tazi.ai**

353 Sacramento St STE 1800  
San Francisco, CA 94111  
+1(408)290-5491  
info@tazi.ai