Rubix ML

- a community videoconference meeting!
- Date and time: On Wednesday, August 5, 2020. Here is time for all time zones. Duration: 2 hours
- Location: https://meet.jit.si/RubixML
- Main goal: for everyone to get to know everyone. Join us!

A high-level machine learning and deep learning library for the php language.

- Developer-friendly API
- 40+ Supervised and Unsupervised Learners
- Support for ETL, Preprocessing, and Cross-validation
- Open Source and Free to use commercially

What

Integrate https://github.com/RubixML/RubixML in Tiki. Rubix ML is a "high-level machine learning and deep learning library"


Both Tiki and Rubix ML are written in PHP, which will facilitate the integration. This is major: Most of the alternatives to Rubix ML are written in Java or Python. We could use them but since they would not be built-in, only a tiny fraction of the Tiki community would have access.

Tiki already has mature data management tools

https://info.tiki.org/Benefits

... Now, in close collaboration with the Rubix ML community, we will add the necessary tools and interfaces to become a complete machine learning platform (managing data, choosing a model, training, evaluation, etc.) accessible to power users, like the rest of Tiki. We will contribute to Rubix ML and make it easier for all other PHP Open Source projects to also integrate with Rubix ML.

Where

- We are coordinating on the Rubix ML chat room, powered by Telegram: https://t.me/RubixML
  - Telegram client apps are Open Source: https://telegram.org/

When

- Tiki22 as an experimental feature.
- Tiki23 as a standard feature
Why

- Permits multiple new features.
  - See "Related" section below for some examples. Many of the these features have been desired for years but we didn't have a clean solution. Both Rubix ML and Tiki have a large feature set and a "one stop shop" philosophy.
  - See various sections at https://github.com/RubixML/RubixML/tree/master/src like Anomaly Detection, Classification, Clustering, etc.

On what types of data?

- On Tiki system data (ex.: logins logs) so will be providing insight for all Tiki instances!
  - Spammy registrations
- On standard features like forums, wiki page, comments, email, etc.
  - Email classification, Spam detection
- On ad hoc data structures made with https://tikitrackers.org/

Who

- Marc (instigator)
- Andrew (Lead dev of Rubix ML) is providing guidance)
- Roberto (developer) will coordinate the project
- Victor (Back-end code) will do initial integration
- Jonny (Front-end code)
- Alain Désilets (advisor)
- Ricardo Melo (advisor)
- Amna Bilal (advisor)
- Simon (junior dev)
- Kevin (junior dev)
- Najia (junior dev)
- Michael I. (tester/requirements for a multilingual project)

How

We'll start with some simple use cases, like reproducing some of the "Project Spotlight" on https://rubixml.com/, but directly within Tiki.

About performance

Performance is very important to train the model. Here is Andrew DalPino, the founder of Rubix ML:

Slides:
https://docs.google.com/presentation/d/1a08XvUzA_9RHtBf55-FOv1XBLgMN7u9dY-Z_E13VXE/edit?usp=sharing
Upcoming meeting

- Mid-July, we'll have an online meeting the Rubix ML and Tiki communities to discuss.

For developers

If you are new to Rubix ML

- Read all the documentation
- Do a quick review of all the code base
- Run at least one of the tutorials: https://docs.rubixml.com/en/latest/#tutorials-example-projects
- Read all the open issues
- Contribute https://github.com/RubixML/RubixML/blob/master/CONTRIBUTING.md

If you are new to Tiki

Reading all the documentation and even a quick scan of all the source code is an unrealistic goal because the project is huge. So just focus at first on Tiki Trackers

1. Join
   - https://wikisuite.chat/ (where Tiki devs hang out)
   - https://tiki.org/login (documentation, forum, bug reports, etc.)
   - Dev Mailing List (where decisions are taken)
2. Read all the content and watch all the videos at https://tikitrackers.org/
3. Install Tiki
   - You can get source from https://gitlab.com/tikiwiki/tiki instead of tarball/zip
4. Explore Tiki features for a few hours
5. Build a simple tracker for yourself
6. Contribute code to Tiki: Git Workflow

Once you know both Rubix ML and Tiki Trackers

- Think about how we could add a graphical user interface (GUI) to Tiki to leverage of Rubix ML.
- Think about how the Rubix ML demos could be handled within Tiki
- Think about how we can have something like MLT without Elasticsearch: https://github.com/RubixML/RubixML/issues/75

Other Ideas For Rubix ML in Tiki

*Actually more like questions - could Rubix ML do things like this?*

- Duplicate content prevention?
  (to display a warning when posting something covered by other content, e.g. in Forums or Trackers)
- Automatic forum moderation - more than just avoiding prevented vocabulary?
- Farming Automation - while a group of drones (100+) are completing an assignment to spray crop protection products (herbicides, pesticides, etc.) on a specific field, several of them will fail. RubixML should recognized when they fail, send out other drones to pick them up, re-organize the workload to account for the sudden changes and send out replacement drones (all controlled by Trackers / List
Execute / Scheduler / etc.)

... Stats

Projects

With Drones


Related

- https://www.linkedin.com/groups/8952251/
- Machine Learning
- Naive Bayes classifier
- Natural language processing
- Auto-Classification
- Optical character recognition
- Text Mining
- IRC QA Mining
- Natural Language Generation
- Farming Automation
- Use Cases for NLP and IR in Tiki
- Follow up about "Is PHP Now Suitable For Machine Learning?"
- Widows and Orphans in mPDF (We will attempt to solve with machine learning)

Machine Learning vs Artificial Intelligence

In French

Video in French about Machine Learning in PHP

Slides:
https://docs.google.com/presentation/d/1XuxgQtlcXuSnLRxSjNPaZBjEPyzg8tj-RzE4wOykyDg/edit?usp=sharing

Other examples

- NextCloud email classification:
  https://github.com/nextcloud/mail/blob/master/lib/Service/Classification/ImportanceClassifier.php