Need help with machine learning? Take the FREE Crash-Course.

Java Machine Learning

by **Jason Brownlee** on July 18, 2014 in **Machine Learning Resources**

Are you a Java programmer and looking to get started or practice machine learning?

Writing programs that make use of machine learning is the best way to learn machine learning. You can write the algorithms yourself from scratch, but you can make a lot more progress if you leverage an existing open source library.

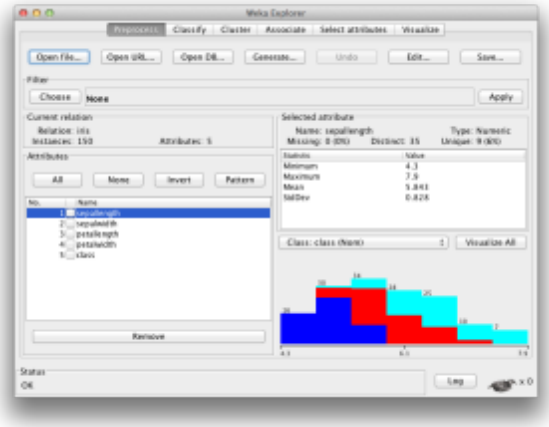
In this post you will discover the major platfo ou can use in Java.

Environments

This section describes Java-based environm e learning. They are called environments beca performing machine learning tasks, but also applications.

Weka

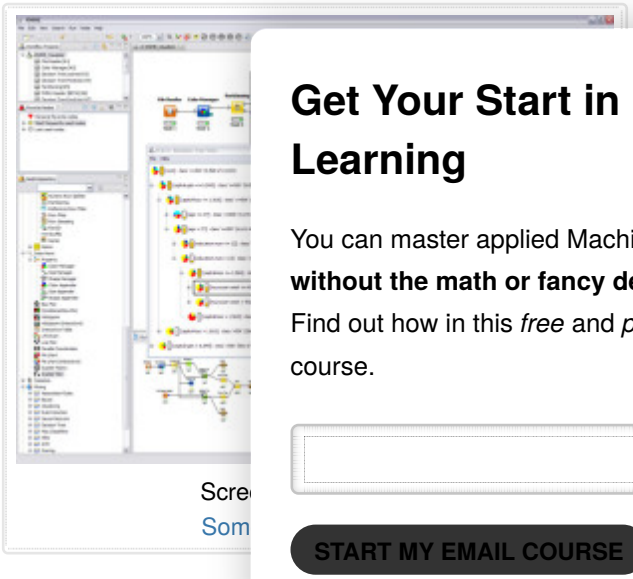
Waikato Environment for Knowledge Analysis (Weka) is a machine learning platform developed by the University of Waikato, New Zealand. It is written in Java and provides a graphical user interface, command line interface and Java API. It is perhaps the most popular Java machine learning library and a great place to start or practice machine learning.



Weka Explorer Interface with the Iris dataset loaded

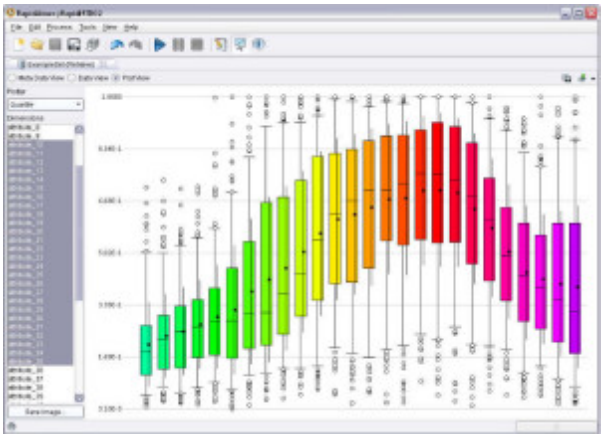
KNIME

The Konstanz Information Miner (KIME) is an analytics and reporting platform developed by Konstanz University, Germany. It was developed with a focus on pharmaceutical research, but has expanded into general business intelligence. It provides a graphical user interface (based on Eclipse) and a Java API.



RapidMiner

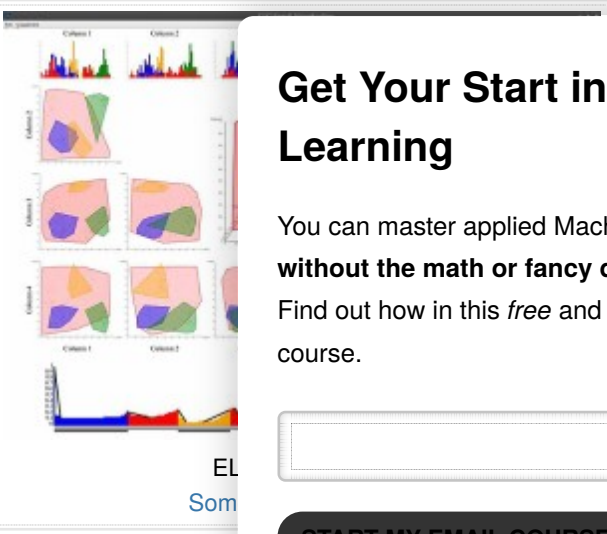
RapidMiner used to be called Yet Another Learning Environment (YALE) and was developed at Technical University of Dortmund, Germany. It provides a GUI and a Java API for developing your own applications. It provides data handling, visualization and modeling with machine learning algorithms.



RapidMiner Screenshot

ELKI

The Environment for DeveLoping KDD-Applications Supported by Index-Structures (ELKI) is a data mining workbench developed in Java by the Ludwig Maximilian University of Munich, Germany. It has a focus on working with data in relational database for tasks such as outlier detection and classification (distance function based methods). It provides a mini GUI, command line interface and Java API.



EL
Som

Libraries

Practically every project listed on this page is/has a library with a Java API, those projects listed in this section only provide a Java API. They are machine learning libraries in the narrow sense.

Java-ML

The Java Machine Learning Library (Java-ML) provides a collection of machine learning algorithms implemented in Java. It provides a standard interface for each algorithm, no UIs and references to the relevant scientific literature for further reading. It includes methods for data manipulation, clustering, feature selection and classification. Note that at the time of writing, the last release was in 2012.

JSAT

The Java Statistical Analysis Tool (JSAT) provides pure Java implementations of standard machine learning algorithms for modest sized problems. The author comments that he developed the library partly as a self-education exercise and partly to get things done. Nevertheless the list of algorithms is impressive. It includes classification, regression, ensemble, clustering and feature selection methods.

Big Data

This section lists Java projects intended for use with Big Data, such as on clusters of machines.

Mahout (Hadoop)

Apache Mahout provides implementations of machine learning algorithms for use on the Apache Hadoop platform (distributed map-reduce). T
classification algorithms and a popular appli
collaborative filtering for recommender syste
run on a single node are also included.

MLlib (Spark)

Apache Machine Learning Library provides i
use on the Apache Spark platform (HDFS, b
the platform support Java, Scala and Pythor
is short, but growing quickly.

MOA

Massive Online Analysis (MOA) is an open source platform designed for data stream mining by University of Waikato, New Zealand. Like Weka (developed at the same place), it provides a GUI, command line interface and Java API. It provides a long list of algorithms wit ha focus on classification and support for outlier detection and addressing concept drift. MOA uses the Advanced Data mining And Machine learning System (ADAMS) for managing workflows also developed at the same place.

SAMOA

Scalable Advanced Massive Online Analysis (SAMOA) is a distributed streaming machine learning framework developed by Yahoo!. It is designed to run on Apache Storm and Apache S4. The system can leverage the algorithms provided by the MOA project for tasks like classification.

Natural Language Processing

This section is dedicated to Java libraries and projects for addressing problems from the subfield of machine learning called Natural Language Processing (NLP).

NLP is not my area, so I'll just point to the key libraries.

- **OpenNLP**: Apache OpenNLP is a toolkit for processing natural language text. It provides methods for NLP tasks such as tokenization, segmentation, and entity extraction.
- **LingPipe**: LingPipe is a toolkit for computational linguistics and includes methods for topic classification, entity extraction, clustering, and sentiment analysis.
- **GATE**: The General Architecture for Text Engineering (GATE) is an open source library for text processing. It provides an array of s
- **MALLET**: Machine Learning for Langua
  natural language processing, document
  information extraction.

Computer Vision

This section lists those libraries for the subfi V).

Again, CV is not my area, so I'll just point to

- **BoofCV**: BoofCV is an open source libr It
  supports features such as image proces
  recognition and image data IO.

Deep Learning

Neural Nets are hot again with the development of deep learning methods and faster hardware. This section lists key Java libraries for working with neural networks and deep learning.

- **Encog**: Encog is a machine learning library that provides algorithms such as SVM, classical neural networks, genetic programming, bayesian networks, HMM and genetic algorithms.
- **Deeplearning4j**: Deeplearning4j is claimed to be a commercial-grade deep learning library written in Java. It is described as being compatible with Hadoop and provides algorithms including Restricted Boltzmann machines, deep-belief networks and Stacked Denoising Autoencoders.

Summary

In this round-up post we have touched on the big name options when selecting a library or platform for machine learning when working in Java.

These are the players and the popular projects, but by no means is this list complete. For example, take a look at this page on MLOSS.org that lists (at the time of writing) 71 Java-based open source machine learning projects. That's a lot and I'm sure there are more on GitHub and SourceForge.

They key is to think hard about your own project and it's requirements. Figure out what you need from a library or platform and then pick and learn a project that best fits your needs.

About Jason Brownlee

Jason Brownlee, Ph.D. is a mac
get results with modern machin
View all posts by Jason Brownle

‹ How to Tune Algorithm Parameters with Scikit-Lear
Applied Machine Learning ion ›

38 Responses to *Java Machine Learning*

Kayode April 8, 2016 at 1:22 pm #

REPLY

This is very interesting. Thanks for sharing. Please have you worked with MOA before? I have some questions to ask on MOA.

Jason Brownlee April 8, 2016 at 1:33 pm #

REPLY

Thanks Kayode.

No sorry, I have not worked with MOA before.

Nemanja May 14, 2016 at 5:15 pm #

REPLY

Thank you very much very educational website

Jason Brownlee May 14, 2016 at 5:37 pm #

REPLY

You're very welcome.

Marcus May 24, 2016 at 7:31 pm #

Just came across SMILE (http://hai

Seems quite complete already. Does anybo

Makes me think about ditching WEKA from

Gaurav Gupta June 13, 2016 at 2:20 a

Thanks for listing the Java based M

missed http://www.h2o.ai/ . Its a java based ML and deep learning APIs and one call/embed it in a JVM based application, please refer – https://www.linkedin.com/pulse/calling-h2o-from-jvm-applications-raymond-peck

Jason Brownlee June 14, 2016 at 8:11 am #

REPLY

Thanks Gaurav.

Generally, I would group H2o as MLaaS more than a Java library.

Swaroop June 25, 2016 at 4:16 am #

REPLY

Great Info. But do you know under MLlib (Spark) section, “Machine” spelling is wrong. Sorry to spam, but couldn’t resist to notify.

Jason Brownlee June 25, 2016 at 5:10 am #

REPLY

Fixed, thanks.

Douglas Arantes July 1, 2016 at 1:39 am #

REPLY

Jason can you add the Smile library to the list?

http://haifengl.github.io/smile/

Jason Brownlee July 1, 2016 at

I’ve not heard of Smile, thanks

Анастасия Ананьева July 15, 2016 at 8:42 pm #

REPLY

We are proud to announce that we moved the website for Java-ML from my personal hosting to Sourceforge .

Yonas September 2, 2016 at 8:46 am #

REPLY

Have you done any kind of programming with the NLP API of the Rapidminer? Or trying to improve any text processing module of the Rapidminer?

Thanks in advance!

Jason Brownlee September 3, 2016 at 6:54 am #

REPLY

Sorry Yonas, I have not.

Gamer Gamer October 23, 2016 at 9:02 am #

REPLY

Hey Jason, i have a java ML project to make for stock market prediction using HMM Model. Which library should i use? How should I proceed? Also, i am totally new to machine learning. Please give me an overview and reply asap!!

Jason Brownlee October 23, 20

Sorry, I am not up to speed on
advice.

For getting started in machine learning,
http://machinelearningmastery.com/start

Nagaraj November 11, 2016 at 4:30 pm #

REPLY

Hi Jason you are website is so good.Thanks for you post.

Jason Brownlee November 12, 2016 at 7:18 am #

REPLY

Thanks Nagaraj.

Nagaraj November 11, 2016 at 4:33 pm #

REPLY

I hope you have good knowledge on machine learning project.Then why don't you start your own product based company.

Jason Brownlee November 12, 2016 at 7:19 am #

REPLY

I have more fun teaching and helping developers get started, thanks Nagaraj.

Michael December 23, 2016 at 1:19 am #

REPLY

Hi Jason,

great post 🙂

Jason Brownlee December 23,

Thanks Michael.

Eyad Farouk January 17, 2017 at 11:3

REPLY

Hello jason,

First of all , I benefit alot from your posts so

Second of all, I want to know if there are specific books to learn machine learning in java. I read your post the other day but the books were all in Python and R. Since I am alredy a java programmer, it would be alot easier to learn the algorithms and build applications in java at the first place instead of learning the algorithms in python and then switching back.
Thanks in advance.

Jason Brownlee January 18, 2017 at 10:15 am #

REPLY

I don't have any material to learn ML in Java at this stage Eyad. Hopefully in the future.

Xiaogang February 13, 2017 at 1:56 pm #

REPLY

Thank you very much for your sharing. Cheers

Jason Brownlee February 14, 2017 at 10:02 am #

REPLY

You're welcome Xiaogang.

Adeel Aslam March 19, 2017 at 2:32 pm #

REPLY

After 2.4 OpenCV now fully suppor

Jason Brownlee March 20, 2017

Thanks for the note Adeel.

Avdhesh Yadav March 22, 2017 at 9:

REPLY

It's great tutorial.What is future in th
auto-generation via spec.

Thanks Avdhesh

Jason Brownlee March 23, 2017 at 8:50 am #

REPLY

Great idea to combine EDA tools with code generation.

gourab April 17, 2017 at 5:16 am #

REPLY

Thank you for much useful article. I want to learn Deep learning/ machine learning where a system should detect an object from video and take certain action. what do you suggest? which one should i learn. i am really new and curious . thanks in advance

Jason Brownlee April 18, 2017 at 8:25 am #

REPLY

Perhaps convolutional neural network that evaluates (classifies) each frame of the video.

erangaz July 29, 2017 at 3:41 pm #

REPLY

Thanx for the great post. I need to create a web navigation agent using ML. Are there any existing algorithms that I can use for predict what is the correct page to navigate?

Jason Brownlee July 30, 2017 a

Sorry, I am not familiar with tha

Perhaps perform some searches on goo

Priya January 3, 2018 at 4:56 pm #

REPLY

Machine learning has basically evo
and computational learning theory. Machine

make high end predictions on data.

Jason Brownlee January 4, 2018 at 8:06 am #

REPLY

Nice.

Blerta March 18, 2018 at 11:38 am #

REPLY

Hi ,
You may find this site http://ramok.tech/machine-learning/ useful for Computer Vision Application implementation in java.

Jason Brownlee March 19, 2018 at 6:04 am #

REPLY

Thanks for sharing.

Leave a Reply

Name (required)

Email (will not be pul

Website

SUBMIT COMMENT

Welcome to Machine Learning Mastery

Hi, I'm Jason Brownlee, Ph.D.
My goal is to make practitioners like YOU awesome at applied machine learning.

Read More

Need More Help? ...take the next step

You can be productive very quickly by taking a systematic approach to machine learning designed for practitioners, not academics.

It is top-down and results-first
with step-by-step tutorials and projects.

No math. No theory. Nothing hidden.

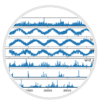
It's time to get serious.

START NOW!

POPULAR

Your First Machine Learning Project in
JUNE 10, 2016

Time Series Prediction with LSTM Recu
JULY 21, 2016



Multivariate Time Series Forecasting wi
AUGUST 14, 2017

How to Setup a Python Environment for
MARCH 13, 2017

Get Your Start in Machine Learning

You can master applied Machine Learning **without the math or fancy degree**.
Find out how in this *free* and *practical* email course.

START MY EMAIL COURSE

ACON

Develop Your First Neural Network in Python With Keras Step-By-Step
MAY 24, 2016

Sequence Classification with LSTM Recurrent Neural Networks in Python with Keras
JULY 26, 2016

Time Series Forecasting with the Long Short-Term Memory Network in Python
APRIL 7, 2017

Regression Tutorial with the Keras Deep Learning Library in Python
JUNE 9, 2016

Multi-Class Classification Tutorial with the Keras Deep Learning Library
JUNE 2, 2016

How to Grid Search Hyperparameters for Deep Learning Models in Python With Keras
AUGUST 9, 2016