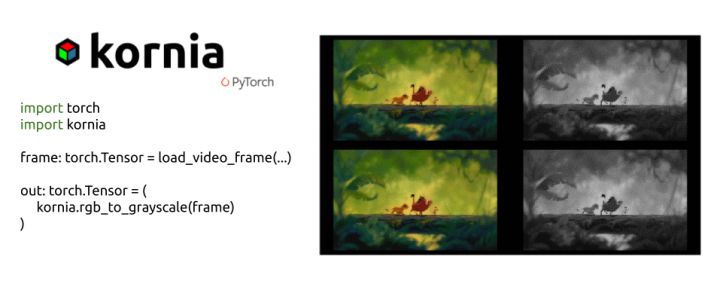
**程序员10 个入门级的机器学习开源项目推荐**

要开始为开源项目做贡献，有一些先决条件：  
1. 学习一门编程语言：由于在开源贡献中你需要编写代码才能参与开发，你需要学习任意一门编程语言。根据项目的需要，在后期学习另一种语言很容易。  
2. 熟悉版本控制系统：这些软件工具有助于将所有更改保存在一个地方，以便在需要时在稍后阶段调用它们。基本上，它们会跟踪在源代码中随时间所做的每一次修改。一些流行的版本控制系统是 Git、Mercurial、CVS 等。其中，Git 是业界最流行和广泛使用的。今天我们来介绍适合AI程序员入门的10个开源机器学习项目。  
1. Caliban  
这是科技巨头Google的机器学习项目。它用于在隔离且可重现的计算环境中开发机器学习研究工作流和笔记本。它解决了一个大问题。 当开发人员在构建数据科学项目时，很多时候很难构建一个可以在现实生活中展示项目的测试环境。因此，Caliban 是解决此问题的潜在解决方案。

项目地址： [google/caliban: Research workflows made easy, locally and in the Cloud. (github.com)](https://github.com/google/caliban)

Caliban 可以轻松地在本地开发任何 ML 模型，在机器上运行代码，然后在云端环境中尝试完全相同的代码以在大型机器上执行。 因此，Dockerized 研究工作流在本地和云端都变得简单。  
2. Kornia

项目地址： [kornia/kornia: Open Source Differentiable Computer Vision Library (github.com)](https://github.com/kornia/kornia)  
Kornia 是 PyTorch 的计算机视觉库。它用于解决一些通用的计算机视觉问题。Kornia 建立在 PyTorch 之上，依靠它的效率和 CPU 能力来计算复杂的函数。



Kornia 是一组用于训练神经网络模型和执行图像变换、图像过滤、边缘检测、对极几何、深度估计等的库。

3. Analytics Zoo

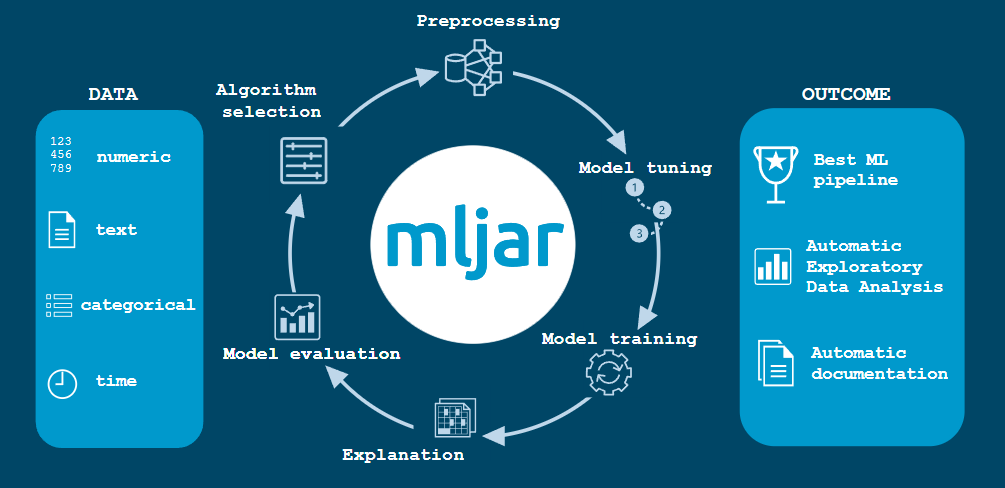
项目地址：[intel-analytics/analytics-zoo: Distributed Tensorflow, Keras and PyTorch on Apache Spark/Flink & Ray (github.com)](https://github.com/intel-analytics/analytics-zoo)  
Analytics Zoo 是一个统一的数据分析和人工智能平台，它将 TensorFlow、Keras、PyTorch、Spark、Flink 和 Ray 程序整合到一个集成管道中。 这可以有效地从笔记本电脑扩展到大型集群以处理大数据的生产。 该项目由 Intel-analytics 维护。  
Analytics Zoo 通过以下方式帮助 AI 解决方案：

* 帮助轻松制作 AI 模型原型。
* 缩放得到有效管理。
* 有助于将自动化流程添加到您的 ML 管道中，例如特征工程、模型选择等。

4. MLJAR 人类自动化机器学习

项目地址：[mljar/mljar-supervised: Python package for AutoML on Tabular Data with Feature Engineering, Hyper-Parameters Tuning, Explanations and Automatic Documentation (github.com)](https://github.com/mljar/mljar-supervised)

MLJAR API ：[mljar/mljar-api-python: A simple python wrapper over MLJAR API. (github.com)](https://github.com/mljar/mljar-api-python)  
Mljar 是一个创建原型模型和部署服务的平台。 为了找到最佳模型，Mljar 搜索不同的算法并执行超参数调整。它通过在云中运行所有计算并最终创建集成模型来提供有趣的快速结果。 然后它会从 AutoML 培训中构建一份报告。 这不是很酷吗？



Mljar 有效地训练用于二元分类、多类分类、回归的模型。  
它提供两种接口：

* 它可以在您的网络浏览器上运行 ML 模型
* 在 Mljar API 上提供 Python 包装器。 [mljar/mljar-api-python: A simple python wrapper over MLJAR API. (github.com)](https://github.com/mljar/mljar-api-python)

从 Mljar 收到的报告包含表格，其中包含有关每个模型分数和训练每个模型所需时间的信息。 性能显示为散点图和箱线图，因此很容易直观地检查哪些算法在所有算法中表现最佳。

5.DeepDetect

项目地址：

[jolibrain/deepdetect: Deep Learning API and Server in C++14 support for Caffe, PyTorch,TensorRT, Dlib, NCNN, Tensorflow, XGBoost and TSNE (github.com)](https://github.com/jolibrain/deepdetect)  
DeepDetect 是一个用 C++ 编写的机器学习 API 和服务器。如果想使用最先进的机器学习算法并希望将它们集成到现有应用程序中，那么 DeepDetect 很适合你。

DeepDetect 支持各种各样的任务，如分类、分割、回归、对象检测、自动编码器。它支持图像、时间序列、文本和更多类型数据的有监督和无监督深度学习。 但是 DeepDetect 依赖于外部机器学习库，例如：

* 深度学习库：Tensorflow、Caffe2、Torch。
* 梯度提升库：XGBoost。
* 使用 T-SNE 进行聚类。

6. Dopamine

项目地址：[google/dopamine: Dopamine is a research framework for fast prototyping of reinforcement learning algorithms. (github.com)](https://github.com/google/dopamine)  
Dopamine 是科技巨头 Google 的一个开源项目。 它是用 Python 编写的。它是一个快速原型强化学习算法的研究框架。  
Dopamine 的设计原则是：  
轻松实验：Dopamine 使新用户可以轻松运行实验。

* 它紧凑而可靠。
* 它还有助于结果的重现性。
* 它很灵活，因此使新用户可以轻松尝试新的研究思路。

7. TensorFlow  
Tensorflow 是 GitHub 上最著名、最受欢迎和最好用的机器学习开源项目之一。它是一个开源软件库，用于使用数据流图进行数值计算。它有一个非常易于使用的 Python 接口，并且没有其他语言中不需要的接口来构建和执行计算图。

TensorFlow 提供稳定的 Python 和 C++ API。 Tensorflow 有一些惊人的用例，例如：

* 在语音/声音识别中
* 文本库应用程序
* 图像识别
* 视频检测
* …还有很多！

提到图像识别与视频检测技术，不得不提目前在各个领域很火的AI+视频技术，将AI检测、智能识别技术融合到各个视频应用场景中，如：安防监控、视频中的人脸检测、人流量统计、危险行为（攀高、摔倒、推搡等）检测识别等。典型的示例如EasyCVR视频融合云服务，具有AI人脸识别、车牌识别、语音对讲、云台控制、声光告警、监控视频分析与数据汇总的能力。

8.PredictionIO

项目地址：[apache/predictionio: PredictionIO, a machine learning server for developers and ML engineers. (github.com)](https://github.com/apache/predictionio)  
它建立在最先进的开源堆栈之上。 该机器学习服务器专为数据科学家设计，可为任何 ML 任务创建预测引擎。 它的一些惊人功能是：

* 它有助于在可定制的生产模板上快速构建和部署引擎作为 Web 服务。
* 部署为 Web 服务后，即可实时响应动态查询。
* 它支持机器学习和数据处理库，如 OpenNLP、Spark MLLib。
* 它还简化了数据基础设施管理

9.Scikit-learn  
它是一个基于 Python 的免费软件机器学习工具库。它提供了用于分类、回归、聚类算法的各种算法，包括随机森林、梯度提升、DBSCAN。

这是建立在必须预先安装的 SciPy 之上，以便可以使用 sci-kit learn。 它还提供以下模型：

* 集成方法
* 特征提取
* 参数调优
* 流形学习
* 特征选择
* 降维

注意：要学习 scikit-learn 遵循文档：  
[https://scikit-learn.org/stable/](https://link.zhihu.com/?target=https%3A//scikit-learn.org/stable/)

10. Pylearn2

项目地址：[lisa-lab/pylearn2: Warning: This project does not have any current developer. See bellow. (github.com)](https://github.com/lisa-lab/pylearn2)  
Pylearn2 是所有 Python 开发人员中最流行的机器学习库。它基于 Theano。你可以使用数学表达式来编写它的插件，而 Theano 则需要优化和稳定。

它有一些很棒的功能，例如：

* 用于训练模型本身的“默认训练算法”
* 模型估计标准
* 分数匹配
* 交叉熵
* 对数似然
* 数据集预处理
* 对比度归一化
* ZCA美白
* 补丁提取（用于实现类卷积算法）