

# Using Machine Learning to Organize the Community - 1

Article



[Renato Banzai](#) · Jul 14, 2020



5m read

[Open Exchange App](#)

## Using Machine Learning to Organize the Community - 1

This is my introduction to a series of posts explaining how to create an end-to-end Machine Learning system.

### Starting with one problem

Our IRIS Development Community has several posts without tags or wrong tagged. As the posts keep growing the organization of each tag and the experience of any community member browsing the subjects tends to decrease.

### First solutions in mind

We can think some usual solutions for this scenario, like:

- Take a volunteer to read all posts and fix the mistakes.
- Pay a company to fix all mistakes.
- Send an email to each post writer to review the texts from past.

### My Solution



Post Tag Classifier 1 Vote in iris-ml-suite!

# Post Tag Classifier Using IRIS + ScikitLearn

Development Community Post URL

Load Post from URL

Text do Predict Tags (you can type or load by URL)

Predict Python SkLearn

Predict IRIS IntegratedML

SkLearn Prediction:

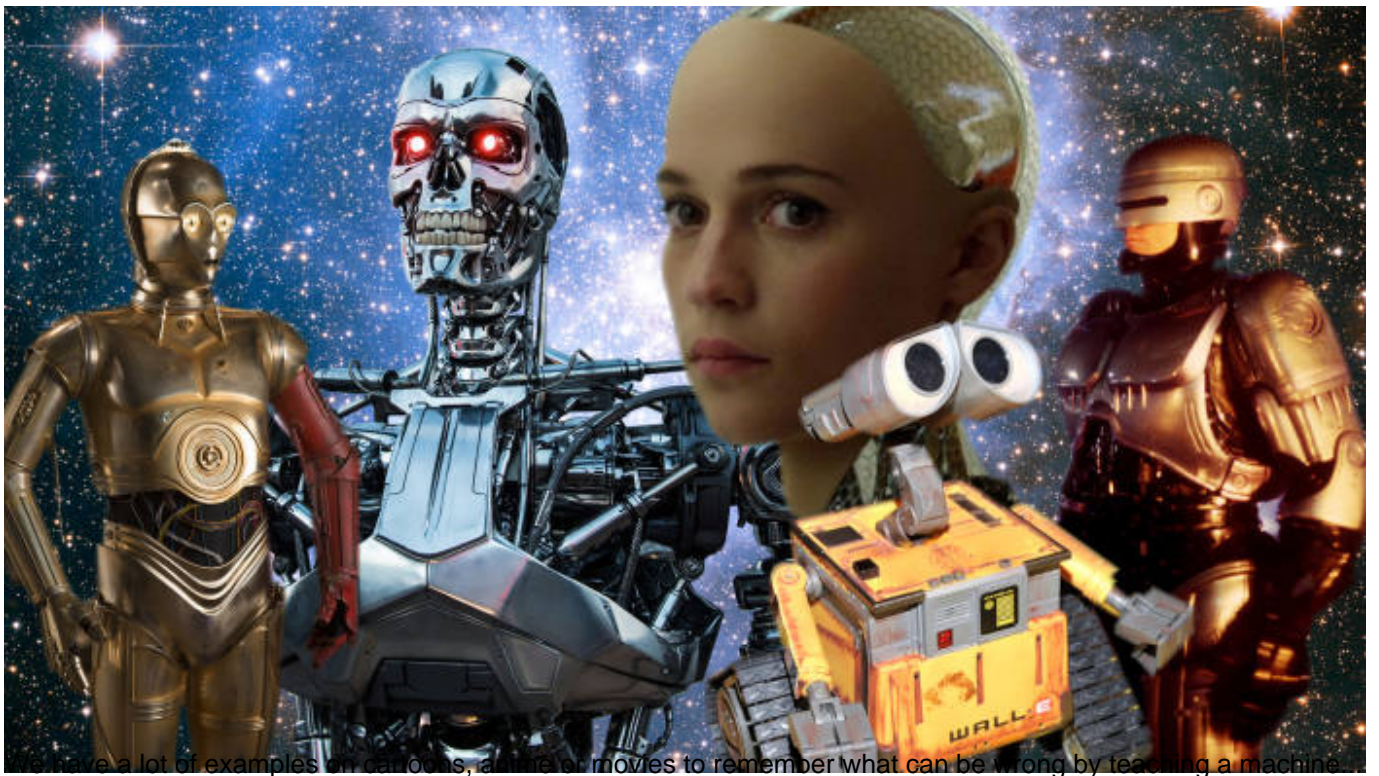
Select...

IRIS IntegratedML Prediction:

Select...



## What if we could teach a machine to do this job?



## Machine Learning

Machine Learning is a very broad topic and I will do my best to explain my vision of the topic. Backing to the problem that

we still need to solve: If we take look at the usual solutions all of then consider interpretation of a text. And how can we teach a machine to read a text, understand the correlation of the text with a tag? First we need to explore the data and take some insights about it.

## Classification? Regression?

When you start to study Machine Learning both of these above terms are always used. But how to know what do you need to go deep?

-Classification: A classification machine learning algorithm predicts discrete values.

-Regression: A regression machine learning algorithm predicts continuous values.

Looking at our problem we need to predict discrete values (all tags exists)

## It's all about data!

All posts data was provided [here](#).

### Post

```
SELECT
  id, Name, Tags, Text
FROM Community.Post
Where
not text is null
order by id
```

id

1946

1951

1956

### Tags

ID

.NET

.NET Experience

ID

AI

API

Now we know how the data looks like. But know the data design isn't enough to create a Machine Learning Model.

### What is a Machine Learning Model?

A machine learning model is a combination of a Machine Learning Algorithm with Data. After combining a technique with data a model can start predicting.

### Accuracy

If you think that ML Models never make mistakes you should understand better the model accuracy. In few words accuracy is how the model perform in predictions. Usually accuracy is expressed in percent like numbers. So someone say "I had created a model with 70% accuracy". This means that for 70% of predictions, the model will predict correctly. The other 30% will go with the wrong prediction.

### NLP - Natural Language Processing

NLP is a field of Machine Learning that works with the ability of a computer to understand and analyse human language. And yes our problem can be solved with NLP.

### Using Machine Learning Algorithms

Most of Machine Learning Algorithms has one thing in common: they use as input **NUMBERS**. Yes I know... this was the most difficult to understand how to create Machine Learning models.

### If all the posts and tags are text how does the model could work?

Good part of the work in a ML Solution is transform the data into something that can be used in a algorithm. This work is called Feature Engineering. In this case is more complicated because the data are unstructured. But a short explanation is\* I transformed each word of text in a unique id represented by a number. SKLearn and other python libs should help you to do this in a easy way.

### Demonstration

I have deployed the trained model as a demo here:

<http://iris-ml-suite.eastus.cloudapp.azure.com/>

### What's next?

In next post I'll show the code and ways to do all the modeling. Don't miss!

### If this article help you or you like the content vote:

This application is at the current contest on open exchange, you can vote in

<https://openexchange.intersystems.com/contest/current> in my application **iris-ml-suite**

---

[Check the related application on InterSystems Open Exchange.](#)

50 1 11 225 [2](#)

Related posts

- Using Machine Learning to Organize the Community - 1
- [Using Machine Learning to Organize the Community - 2](#)

[Show all](#)

**Source URL:** <https://community.intersystems.com/post/using-machine-learning-organize-community-1>