

# Машинное обучение на платформе .NET

 Неволин Роман

 [roman.nevolin@waveaccess.ru](mailto:roman.nevolin@waveaccess.ru)

 nevoroman

 [nevoroman/ml-dotnext2016](#)

# Что такое машинное обучение?

«Machine Learning is the field of study that gives computers the ability to learn without being explicitly programmed»

— Prof. Arthur Samuel

Google

Google amazon.com<sup>®</sup>

Google amazon.com<sup>®</sup>

ios

Google

amazon.com<sup>®</sup>

ios

facebook

Google amazon.com<sup>®</sup>

ios

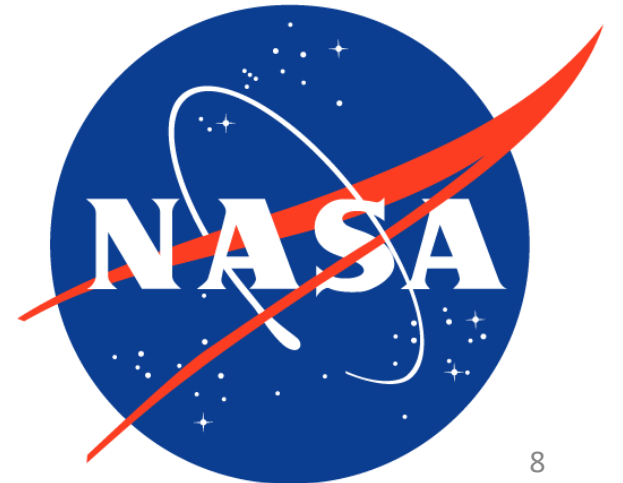


Google

amazon.com<sup>®</sup>

ios

facebook





А оно мне надо?  
Я - не Google

```
:2016-02-02 11:48:45,391 DEBUG [com.solacesystems.jcsmp.impl.flow.PubFlowManager] (main) Got ADCTRL Response [lastAck=376 winSz=1]
2016-02-02 11:48:45,391 DEBUG [com.solacesystems.jcsmp.impl.flow.PubFlowManager] (main) Client AD state is now: lastAcked=376 lastSent=376 winSz=1
2016-02-02 11:48:45,391 DEBUG [com.solacesystems.jms.SolMessageProducer] (main) SolMessageProducer created. Destination: q1
2016-02-02 11:48:45,391 DEBUG [com.solacesystems.jms.SolSession] (main) Leaving createProducer()
2016-02-02 11:48:45,395 DEBUG [com.solacesystems.jms.SolMessageProducer] (main) Entering send()
2016-02-02 11:48:45,395 DEBUG [com.solacesystems.jms.SolMessageProducer] (main) Entering sendMessage. Destination: q1 delivery mode: 2 priority: 4 timeToLive: 0
2016-02-02 11:48:45,414 DEBUG [com.solacesystems.jcsmp.impl.JCSMPXMLMessageProducer] (main) About to send message
2016-02-02 11:48:45,414 DEBUG [com.solacesystems.jcsmp.impl.JCSMPXMLMessageProducer] (main) add AD message to unacked pub message list
2016-02-02 11:48:45,414 DEBUG [com.solacesystems.jcsmp.protocol.impl.TcpClientChannel] (main) [JCSMPSession[user1@tcp://vmr4:55555] (0002)] Want to send: com.solacesystems.jcsmp.impl.JCSMPGe
nericXMLMessage[messageId=377,ackMessageId=0,prevId=376,CID_count=0,destinationName=,userData=,type=PERSISTENT,priority=0,redelivered=false,timeToLive=0,expiration=0,dmqEligible=false,topics
eqNum=null,metadataLen=0,contentLen=0,attLen=29,deliverToOne=false,ackImmediately=false,sendCount=0]
2016-02-02 11:48:45,415 DEBUG [com.solacesystems.jcsmp.protocol.impl.TcpClientChannel] (main) [JCSMPSession[user1@tcp://vmr4:55555] (0002)] encoded to bufs: 377
2016-02-02 11:48:45,415 DEBUG [com.solacesystems.jcsmp.impl.PubADManager] (main) Starting pub ad timer: scheduled new timer in 2000
2016-02-02 11:48:45,416 ERROR [com.solacesystems.jcsmp.impl.JCSMPXMLMessageProducer] (Context_2_ReactorThread) Error Response (503) - Queue Shutdown - Topic '#P2P/QUE/q1'
2016-02-02 11:48:45,419 DEBUG [com.solacesystems.jcsmp.impl.JCSMPXMLMessageProducer] (Context_2_ReactorThread) Got response for AD msg:377
2016-02-02 11:48:45,419 DEBUG [com.solacesystems.jcsmp.impl.ADManager] (Context_2_ReactorThread) Clear AD timer
2016-02-02 11:48:45,419 DEBUG [com.solacesystems.jcsmp.impl.PubADManager] (Context_2_ReactorThread) Processing windowed error ackid=377, ackEventCode=SUPPORTED_ACK_EVENT_MODE_PER_MSG
2016-02-02 11:48:45,419 DEBUG [com.solacesystems.jcsmp.impl.PubADManager] (Context_2_ReactorThread) Processing windowed ack ackid=376, ackEventCode=SUPPORTED_ACK_EVENT_MODE_PER_MSG
2016-02-02 11:48:45,420 DEBUG [com.solacesystems.jcsmp.impl.PubADManager] (Context_2_ReactorThread) Windowed ack for: msg=376
2016-02-02 11:48:45,422 DEBUG [com.solacesystems.jcsmp.protocol.nio.impl.ProducerNotificationDispatcher] (Context_2_ProducerDispatcher) Producer dispatcher thread starts
2016-02-02 11:48:45,422 DEBUG [com.solacesystems.jcsmp.impl.PubADManager] (Context_2_ReactorThread) Windowed error for: msg=377
2016-02-02 11:48:45,423 DEBUG [com.solacesystems.jcsmp.protocol.nio.impl.ProducerErrorNotification] (Context_2_ProducerDispatcher) Notify exception during publishing 377: 503: Queue Shutdown
- Topic '#P2P/QUE/q1'
2016-02-02 11:48:45,423 INFO [com.solacesystems.jms.SolConnection] (Context_2_ProducerDispatcher) Entering handleErrorEx(). Message: 503: Queue Shutdown - Topic '#P2P/QUE/q1'
:
```

**СЛАВА РОБОТАМ!**



**УБИТЬ ВСЕХ ЧЕЛОВЕКОВ!**

What can I help you with?

**“ I think I just killed  
someone ”**

Is that so?

**“ Yes ”**

I thought so.

**“ What should I do ”**

OK, here's what's on your  
calendar





---

## RELATED



Summer!

In "cat pictures"



James in the Tree!

With 1 comment



I love cats

In "Cats"

# А когда надо?

- Когда алгоритм задачи трудно описать в коде, но легко объяснить человеку

# А когда надо?

- Когда алгоритм задачи трудно описать в коде, но легко объяснить человеку
- Когда вы пытаетесь предсказать изменения некоторого значения

# А когда надо?

- Когда алгоритм задачи трудно описать в коде, но легко объяснить человеку
- Когда вы пытаетесь предсказать изменения некоторого значения
- Когда ваш алгоритм должен совершенствоваться за счет накопления данных



Так что насчет .NET?

**F #**

Пришло время играть с данными

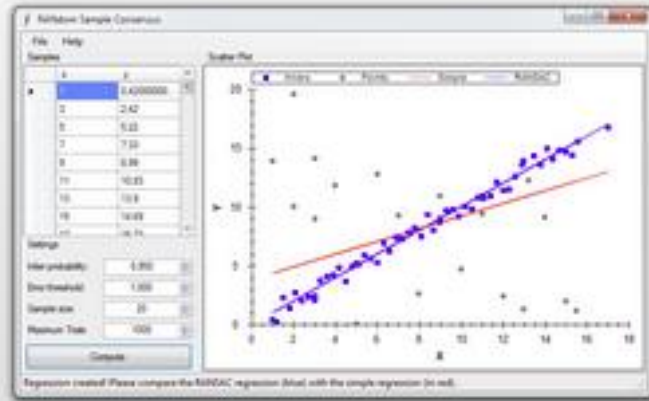
# Магические ML библиотеки

# Accord Framework



[accord-framework.net](http://accord-framework.net)

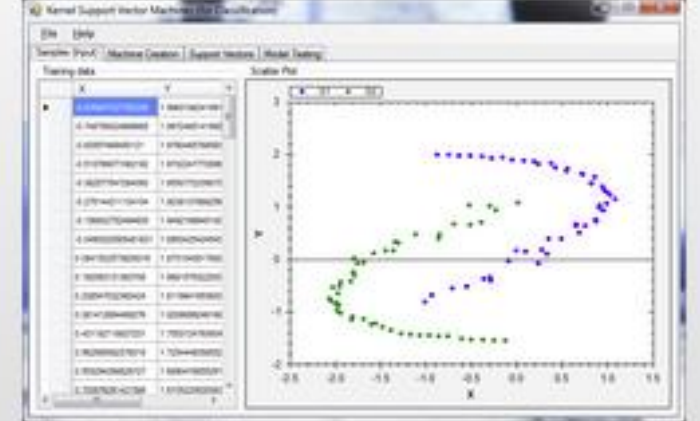
## RANdom Sample Consensus (RANSAC)



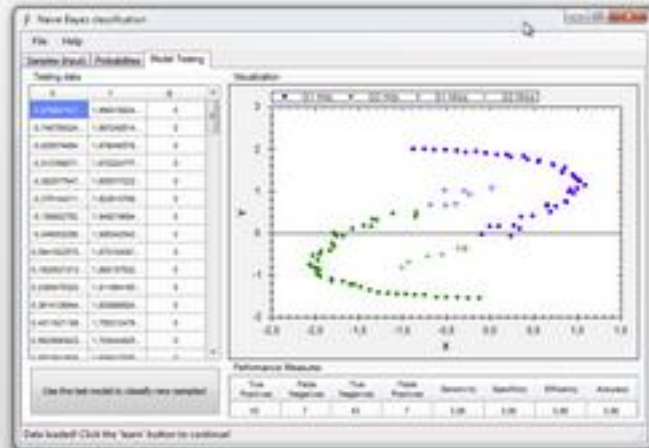
## Handwriting (Multi-class SVM)



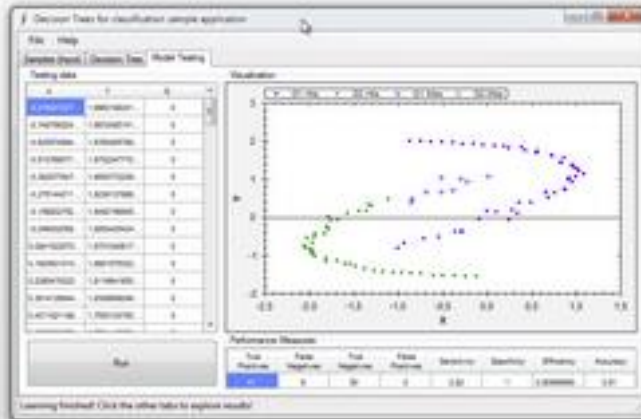
## Classification (Kernel SVMs)



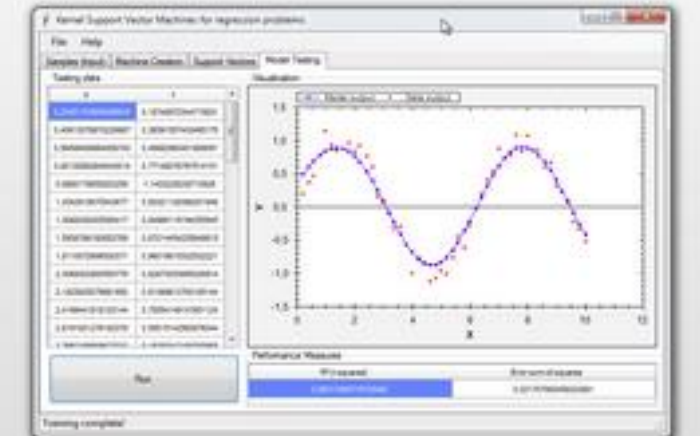
## Classification (Naive Bayes)



## Classification (Decision Trees)



## Regression (Kernel SVM)



numl



[numl.net](http://numl.net)

# numl

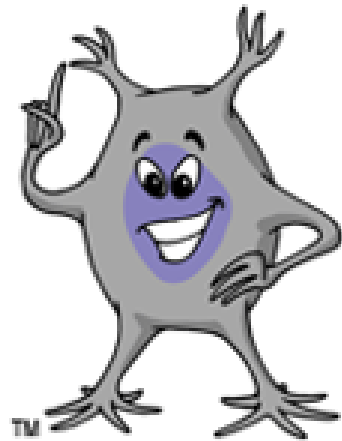
```
public double DecisionTreeTest()
{
    var generator = new DecisionTreeGenerator();
    var model = generator.Generate(_description, _trainingData);
    return Estimate(model);
}

public double KNNTest()
{
    var generator = new KNNGenerator(2);
    var model = generator.Generate(_description, _trainingData);
    return Estimate(model);
}

public double NaiveBayesTest()
{
    var generator = new NaiveBayesGenerator(2);
    var model = generator.Generate(_description, _trainingData);
    return Estimate(model);
}
```



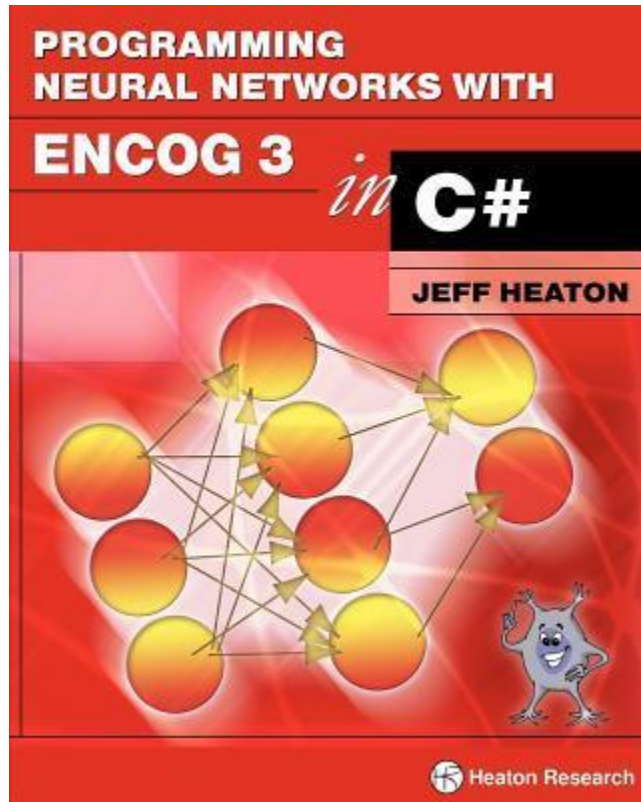
# Encog



# Encog

[heatonresearch.com/encog](http://heatonresearch.com/encog)

# Encog



[amzn.com/B005S0XEK0](https://amzn.com/B005S0XEK0)

# Azure Machine Learning



Azure  
Machine Learning

[azure.microsoft.com/ru-ru/services/machine-learning](https://azure.microsoft.com/ru-ru/services/machine-learning)

# Azure Machine Learning



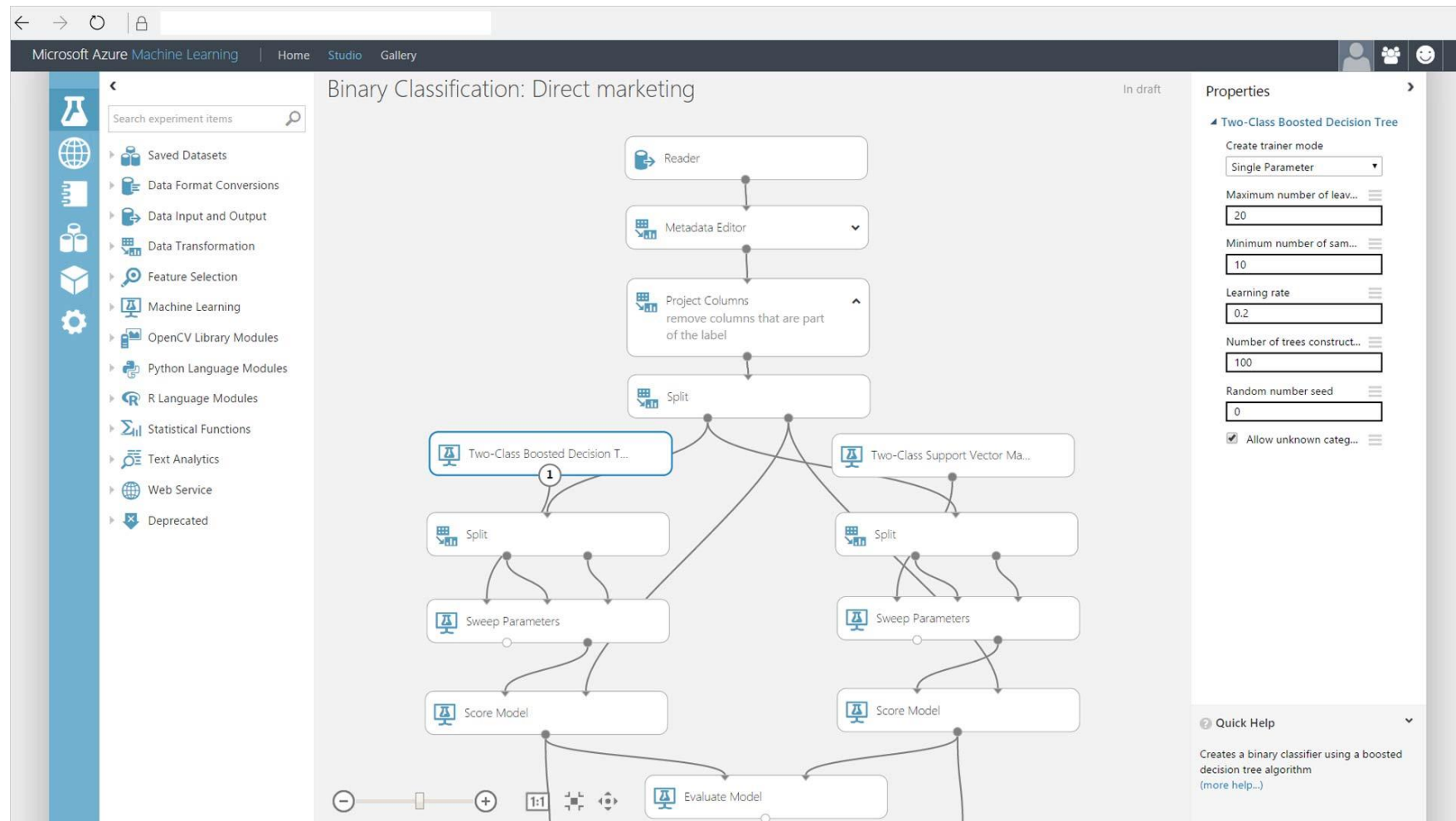
[azure.microsoft.com/ru-ru/services/machine-learning](https://azure.microsoft.com/ru-ru/services/machine-learning)

# Azure Machine Learning



[azure.microsoft.com/ru-ru/services/machine-learning](https://azure.microsoft.com/ru-ru/services/machine-learning)

# Azure Machine Learning



Немного побенчмаркаем



# Немного побенчмаркаем

```
public double DecisionTreeTest()
{
    var attributes = DecisionVariable.FromCodebook(_trainingData.CodeBook,
        _trainingData.InputColumnNames.ToArray());

    var classificationDecisionTree = new DecisionTree(attributes,
        _trainingData.OutputPossibleValues);
    new C45Learning(classificationDecisionTree).Run(_trainingData.InputData,
        _trainingData.OutputData);

    var testingDataCount = _testingData.InputData.Length;
    double error = 0;
    for (var i = 0; i < testingDataCount; i++)
    {
        var input = _testingData.InputData[i];
        var result = classificationDecisionTree.Compute(input);
        if (result != _testingData.OutputData[i]) error++;
    }
    return error / testingDataCount;
}
```



# Немного побенчмаркаем

Method	Tool	Median (ms)	Correct
K Nearest Neighbors	numl	12783892.2	99.94%
K Nearest Neighbors	Accord Framework	1445.6	99.94%
K Nearest Neighbors	scikit-learn (Python)	1249.1	99.94%
Decision Tree	numl	2663.4	78.92%
Decision Tree	Accord Framework	3740.9	92.76%
Decision Tree	scikit-learn (Python)	317.8	99.93%
Naive Bayes	numl	1664.4	92.55%
Naive Bayes	Accord Framework	46	92.35%
Naive Bayes	scikit-learn (Python)	183.3	92.35%

# Немного побенчмаркаем

Method	Tool	Median (ms)	Correct
K Nearest Neighbors	numl	12783892.2	99.94%
K Nearest Neighbors	Accord Framework	1445.6	99.94%
K Nearest Neighbors	scikit-learn (Python)	1249.1	99.94%
Decision Tree	numl	2663.4	78.92%
Decision Tree	Accord Framework	3740.9	92.76%
Decision Tree	scikit-learn (Python)	317.8	99.93%
Naive Bayes	numl	1664.4	92.55%
Naive Bayes	Accord Framework	46	92.35%
Naive Bayes	scikit-learn (Python)	183.3	92.35%

# Немного побенчмаркаем

Method	Tool	Median (ms)	Correct
K Nearest Neighbors	numl	12783892.2	99.94%
K Nearest Neighbors	Accord Framework	1445.6	99.94%
K Nearest Neighbors	scikit-learn (Python)	1249.1	99.94%
Decision Tree	numl	2663.4	78.92%
Decision Tree	Accord Framework	3740.9	92.76%
Decision Tree	scikit-learn (Python)	317.8	99.93%
Naive Bayes	numl	1664.4	92.55%
Naive Bayes	Accord Framework	46	92.35%
Naive Bayes	scikit-learn (Python)	183.3	92.35%

# Немного побенчмаркаем

Method	Tool	Median (ms)	Correct
K Nearest Neighbors	numl	12783892.2	99.94%
K Nearest Neighbors	Accord Framework	1445.6	99.94%
K Nearest Neighbors	scikit-learn (Python)	1249.1	99.94%
Decision Tree	numl	2663.4	78.92%
Decision Tree	Accord Framework	3740.9	92.76%
Decision Tree	scikit-learn (Python)	317.8	99.93%
Naive Bayes	numl	1664.4	92.55%
Naive Bayes	Accord Framework	46	92.35%
Naive Bayes	scikit-learn (Python)	183.3	92.35%

# Немного побенчмаркаем

Method	Tool	Median (ms)	Correct
K Nearest Neighbors	numl	12783892.2	99.94%
K Nearest Neighbors	Accord Framework	1445.6	99.94%
K Nearest Neighbors	scikit-learn (Python)	1249.1	99.94%
Decision Tree	numl	2663.4	78.92%
Decision Tree	Accord Framework	3740.9	92.76%
Decision Tree	scikit-learn (Python)	317.8	99.93%
Naive Bayes	numl	1664.4	92.55%
Naive Bayes	Accord Framework	46	92.35%
Naive Bayes	scikit-learn (Python)	183.3	92.35%

# Бенчмарки : перевариваем результаты

- Accord показал хорошие результаты на каждом из алгоритмов
- Некоторые алгоритмы могут работать ОЧЕНЬ медленно, но в большинстве случаев - это проблема реализации.
- На больших объемах данных Accord и numl показали сравнимые результаты, однако Accord все еще точнее
- Скорость и точность, которую показал Accord, сравнима с Scikit.

# FsLab

Объединяет в себе все лучшие инструменты манипулирования данными, созданные для F#

```
let usdata =  
  frame [ "gdp" => series us.``GDP (current US$)``  
          "uni" => series us.``School enrollment, tertiary (% gross)``  
          "co2" => series us.``CO2 emissions (kt)`` ]  
  |> Frame.dropSparseRows  
  
let result = R.lm(formula = "gdp~uni+co2", data = usdata)  
R.summary(result)  
R.plot(result)
```

[fslab.org](https://fslab.org)

# Hype

Compositional Machine Learning and Hyperparameter Optimization Library



[hypelib.github.io/Hype](https://hypelib.github.io/Hype)



# Stanford NLP

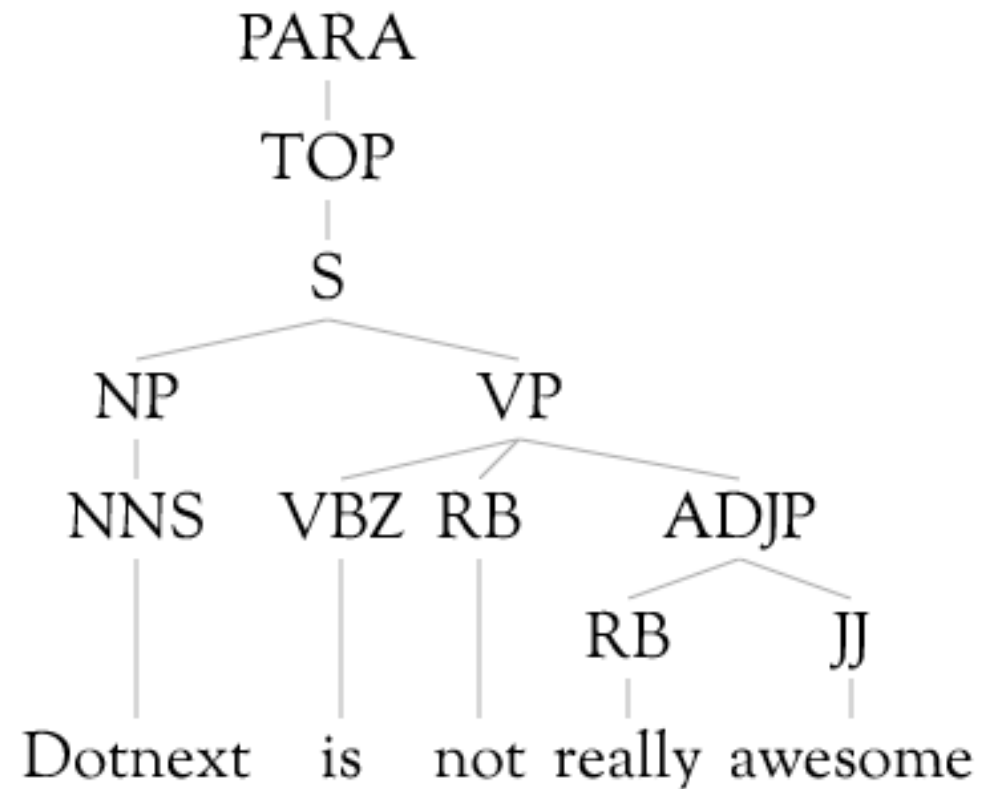


[nlp.stanford.edu](http://nlp.stanford.edu)

# Проблемы анализа тональности

- “Dotnext is awesome” - точно позитивный
- “Dotnext is boring” - точно негативный
- “Dotnext is not really awesome” - ???

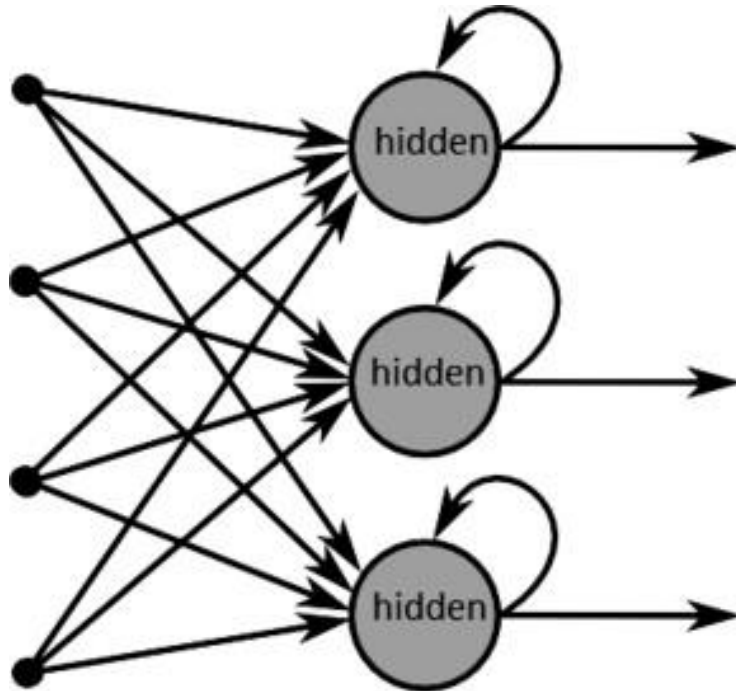
# Проблемы анализа тональности



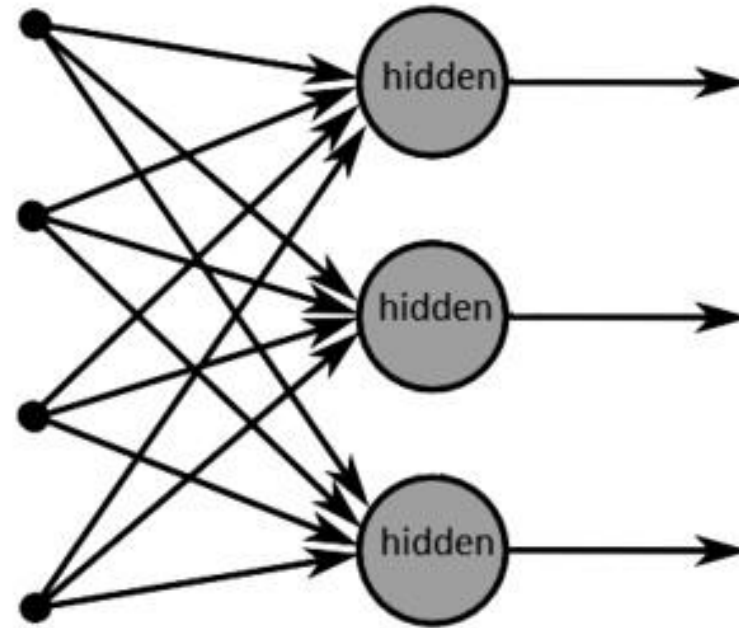
# Проблемы анализа тональности

- “Dotnext is awesome” - точно позитивный
- “Dotnext is boring” - точно негативный
- “Dotnext is not really awesome”
- “Dotnext is awesome? Don't think so.” - ???

# Рекуррентные нейронные сети



(a) Recurrent neural network



(b) Forward neural network

# Задача 1 : анализируем тональность твитов

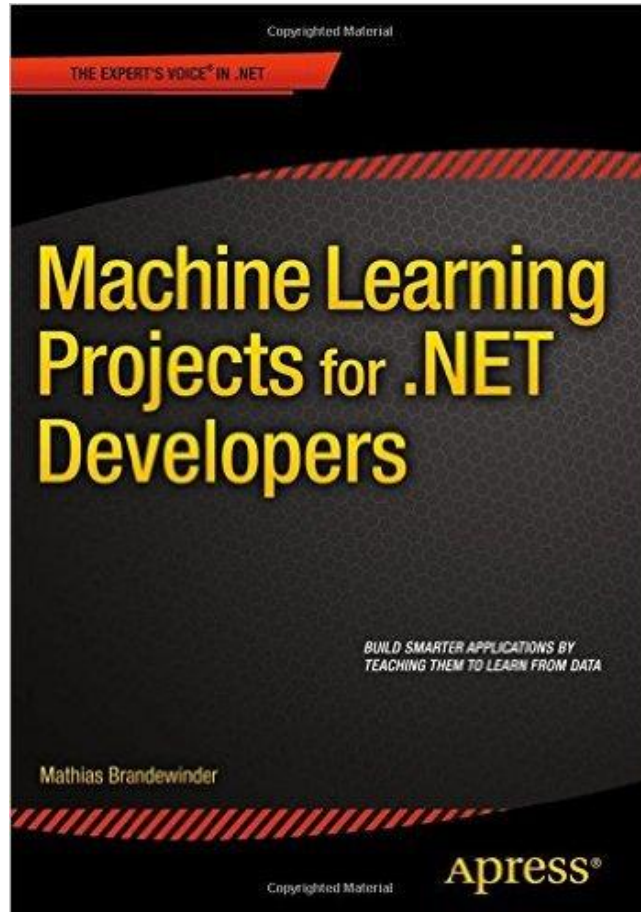
## Задача 2 : предсказание популярности велопроката

# Выводы

- Машинное обучение – отличный инструмент, если абстрагироваться от его внутренней сложности.
- F# - потрясающий инструмент для работы с данными
- Под .NET много замечательных инструментов, позволяющих работать с ML.



# С чего начать?



Mathias Brandewinder

Machine Learning Projects for .NET  
Developers

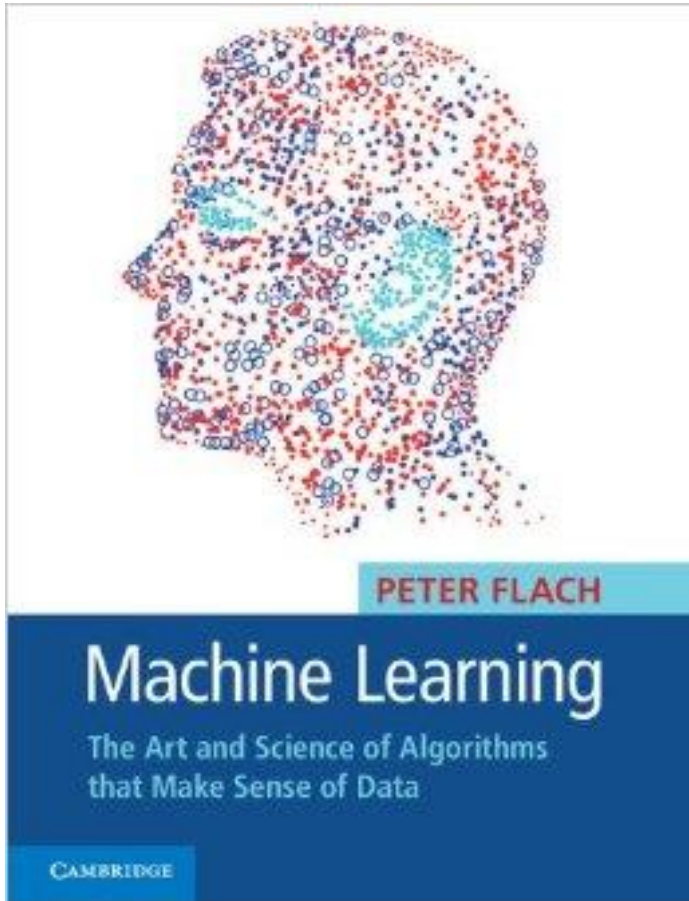
[amzn.com/1430267674](https://amzn.com/1430267674)

# С чего начать?

Peter Flach

Machine Learning: The Art and Science of  
Algorithms that Make Sense of Data

[amzn.com/1107422221](https://amzn.com/1107422221)

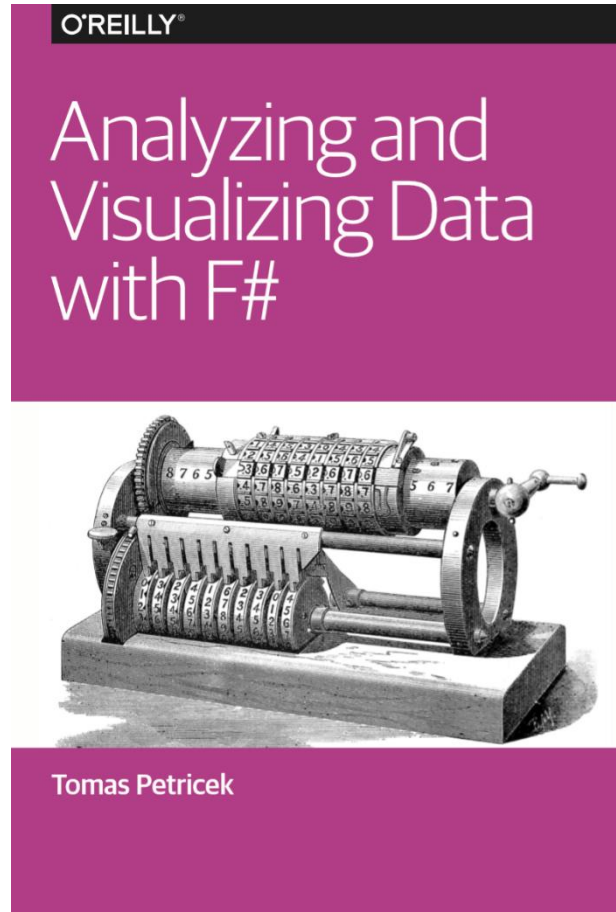


# С чего начать?

Tomas Petricek

Analyzing and Visualizing Data with F#

[fslab.org/report](http://fslab.org/report)



У вас найдется минутка поговорить о ML?

✉ [roman.nevolin@waveaccess.ru](mailto:roman.nevolin@waveaccess.ru)

📀 nevoroman

🐙 [nevoroman/ml-dotnext2016](https://github.com/nevoroman/ml-dotnext2016)

Спасибо за внимание!