THE
DEVELOPER'S
CONFERENCE

Building a Text Classifier with NaturalLanguage and CoreML

Alan Paiva

JULHO | 2019

@ajeferson

ALAN PAIVA

iOS Software Engineer at **ArcTouch,** Reactive Programming and **Machine Learning** enthusiast, blog writer. Power metal **headbanger** while doing all of that.


We're hiring!
arctouch.com/careers/brazil

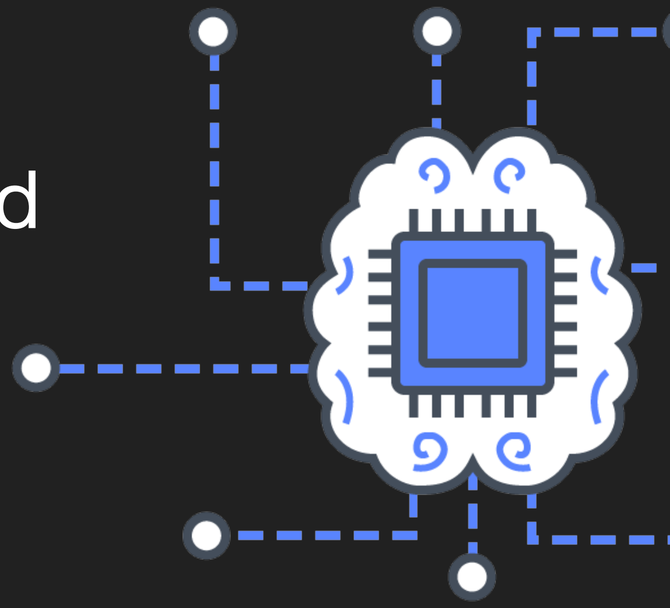
AGENDA

- Intro
- NaturalLanguage Framework
- Machine Learning
- Word Embeddings

Intro

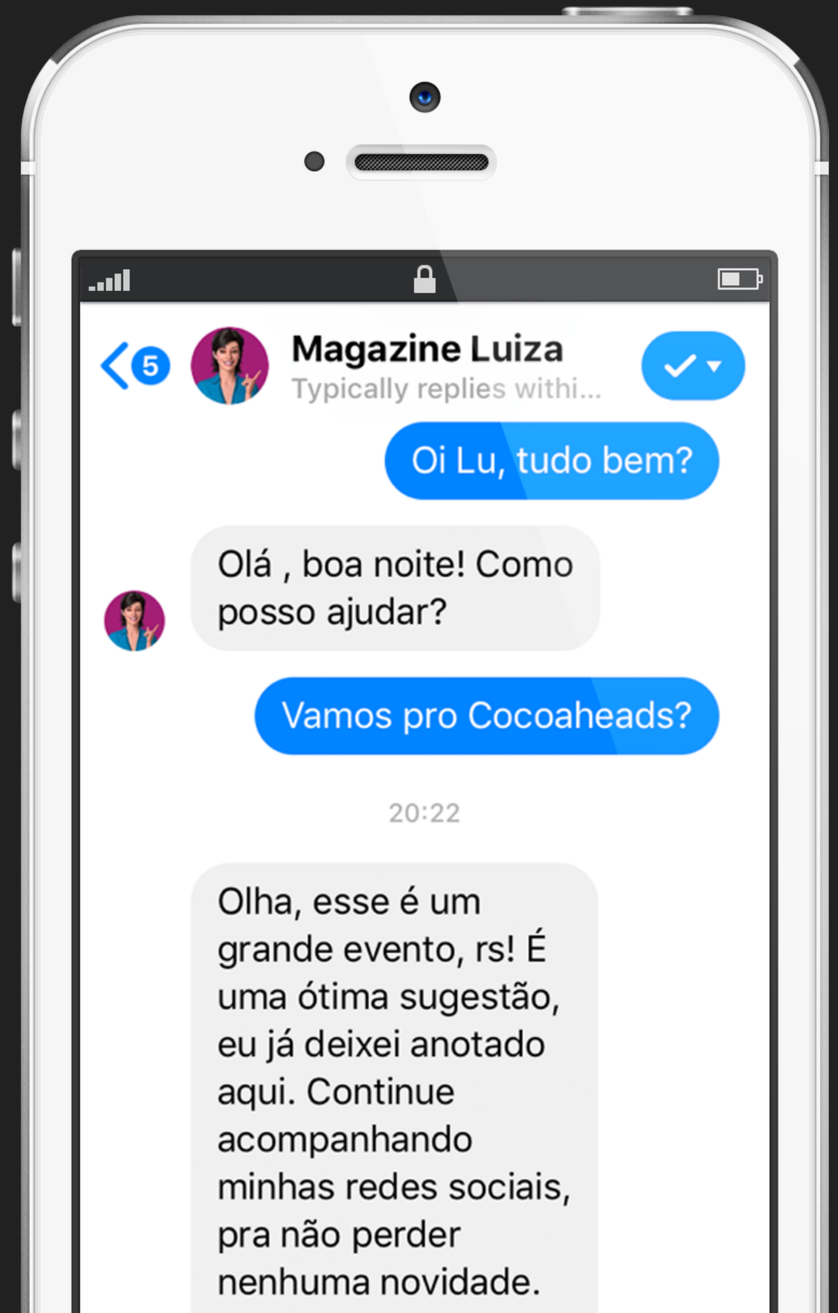
NLP

- In a wide sense, it refers to **any kind of natural language manipulation** performed by computers
- Field of **Machine Learning** that seeks for understanding human language



APPLICATIONS

- Text **Categorization**
- Machine **Translation**
- **Chat Bots**
- Virtual **Assistants**
- **Sentiment** Analysis
- **SPAM** Analysis



Natural Language Framework

NATURAL LANGUAGE FRAMEWORK

Built-in Features

- Language **Identification**
- **Tokenization**
- **POS-Tagging**
- Named **Entity Recognition**

NATURAL LANGUAGE FRAMEWORK

- Integration with CreateML and CoreML
- Customizable Models
  - Text Classification
  - Word Tagging

Text Classification

TEXT	TOPIC
However, the site earlier announced that it was considering introducing restrictions on live-streaming in the wake of the Christchurch attacks. On Thursday, it also said that it would ban white nationalism and separatism from the site.	**NEWS**
*And God said, “Let there be light,” and there was light. **4** God saw that the light was good, and he separated the light from the darkness. 5 God called the light “day,” and the darkness he called “night.” And there was evening, and there was morning—the first day.*	**RELIGION**

Machine Learning

Supervised Learning

TRAINING

TRAINING SET → **FEATURE** EXTRACTION → **FITTING** → MODEL **(KNOWLEDGE)**

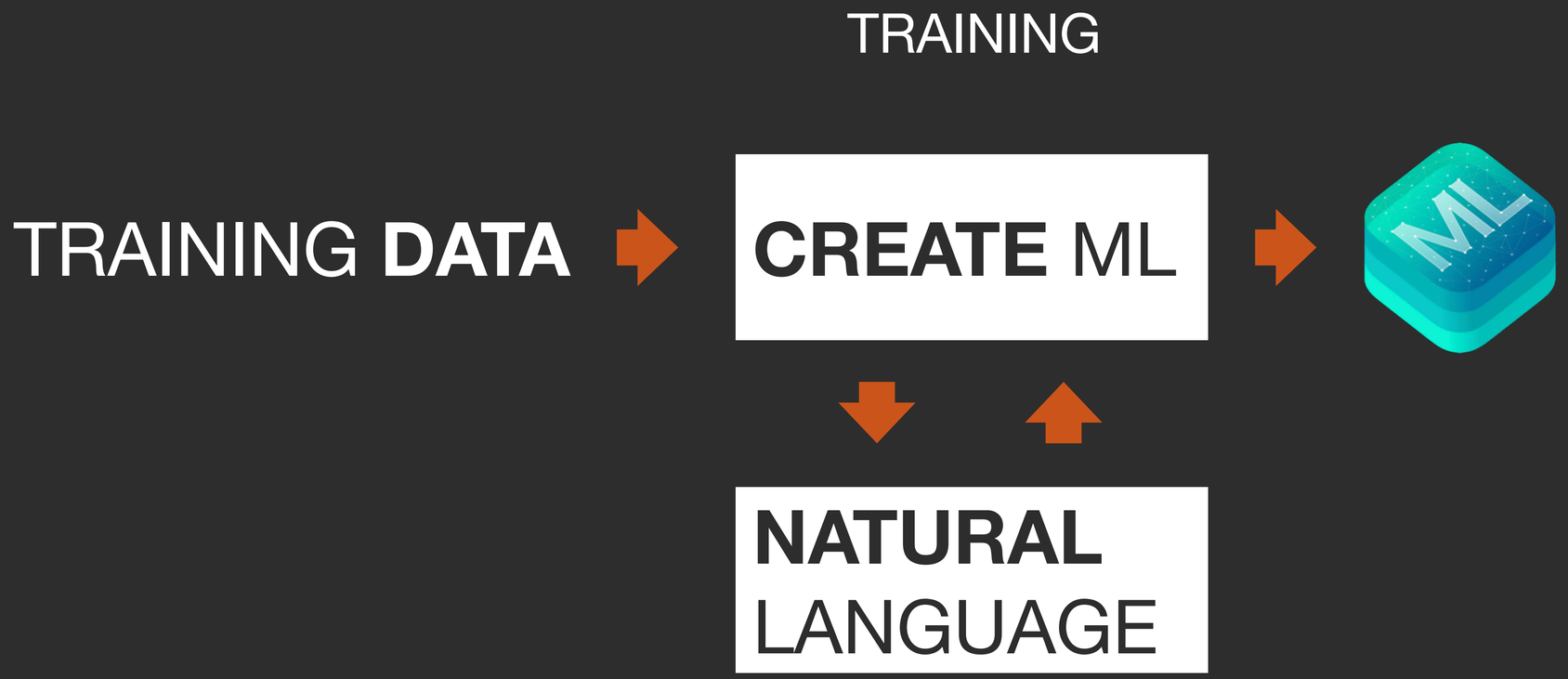
TESTING

Supervised Learning

PREDICTING

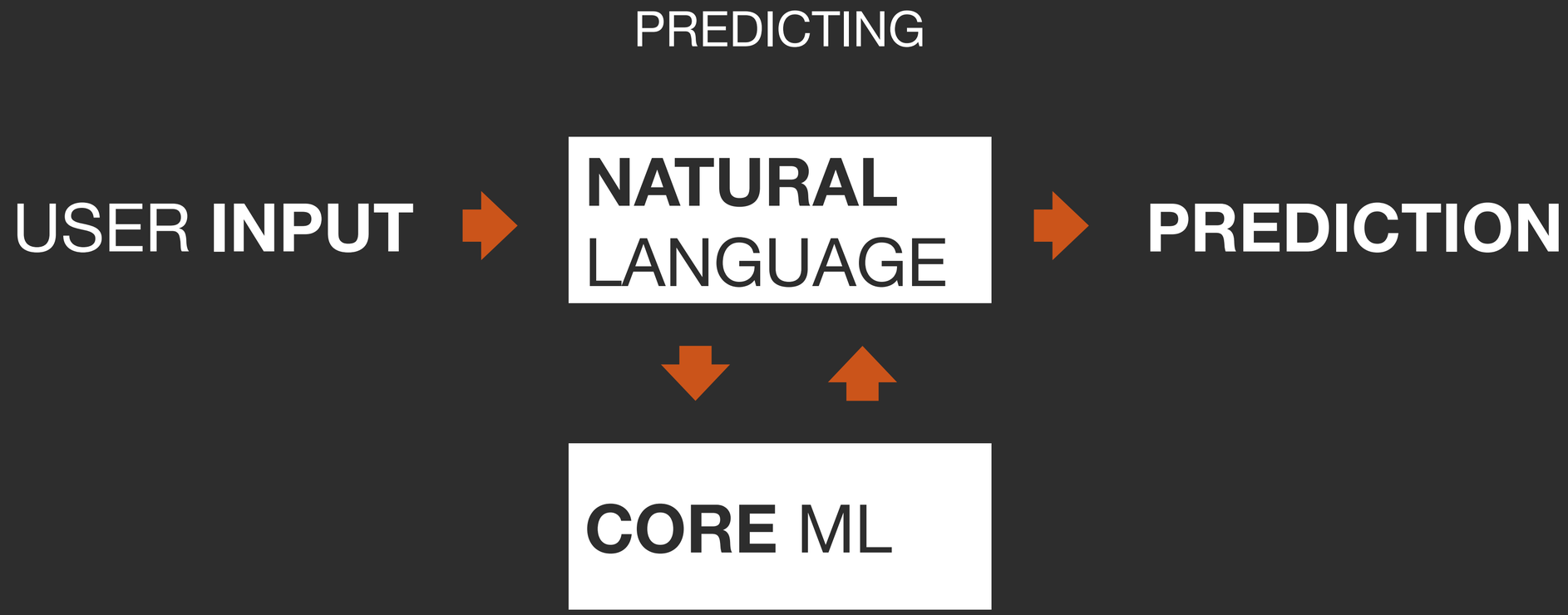
USER **DATA** → **FEATURE** EXTRACTION → **INFERENCE** → **PREDICTION**

Supervised Learning



Supervised Learning

NaturalLanguage Framework



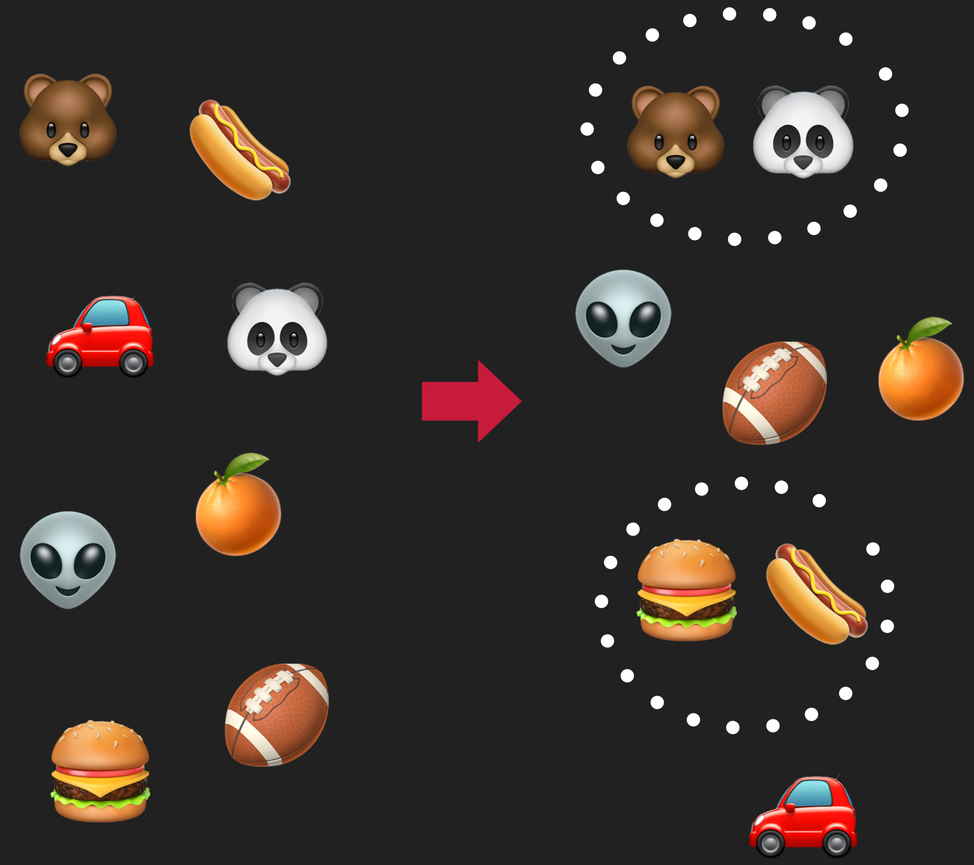
DEMO


TURNIP 2000
INITIATING....WAIT...WAIT A BIT MORE... ALMOST
FUNK SUBROUTINE V2.09
WELCOME TO FUNKTOWN, GROOVER
DEFINE (FUNK PARAMTERS: SUPER CRUNKY
WORKING.....
(CRUNKYness SET TO SETTING: SUPER(SUPER CRUNKY)
C:/LOVE_THE_WAY_YOU_MOVE.MP3 -LOAD TO EXT. DEVICE
WRITTEN SUCCESSFULLY TO EXT. DEVICE
C:/DROP_DA_BEAT.EXE -P(LOVE THE_WAY_YOU_MOVE.MP3)
BEAT DROPPING.....
BEAT SUCCESSFULLY DROPPED. IS IT FRESH? Y/N
Y
TIGHT. INITIATE FLY SOVIET DANCE KICKS TESTING
PHASE? Y/N
Y
INITIATING....PLEASE WAIT.....
.....
SUCCESSFULLY INITIATED.
C:/BASH
BASH 3.20 LX -ART
FUNK DROPS MAD BEATS FLY KICKS
DANCE POPS LOCKS CRUNK
BASH 3.2: CD DANCE

Word Embeddings

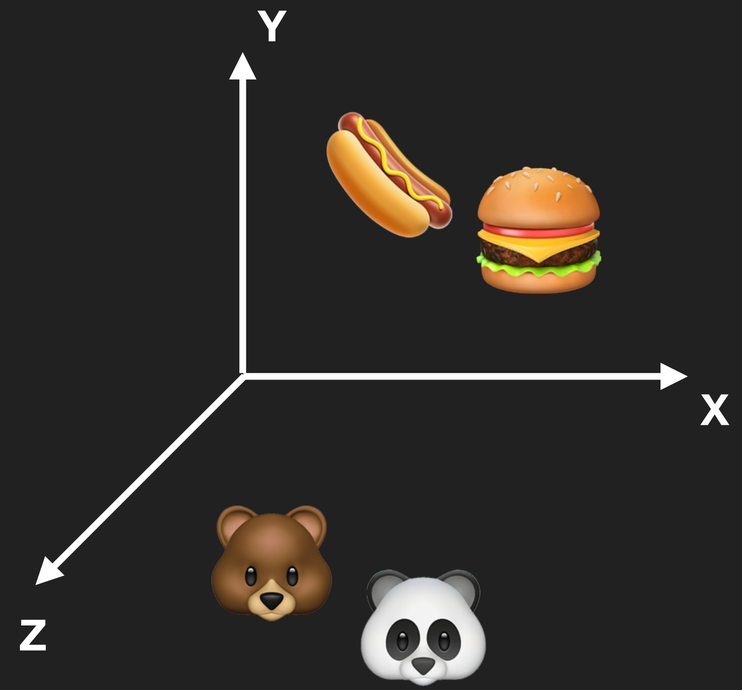
WORD EMBEDDINGS

- Groups words that are **similar in meaning**
- Map strings to **vectors** in a n-dimensional space



WORD EMBEDDINGS

🐻	9.874	6.409	7.329
🐼	9.539	6.087	6.997
...	...	...	...
🍔	-2.341	-6.455	-1.532
🌭	-2.312	-6.502	-1.704



WORD EMBEDDINGS

```
guard let embedding = NLEmbedding.wordEmbedding(for: .english) else { return }

let dimensions = embedding.dimension

guard let computerVector = embedding.vector(for: "computer") else { return }

let laptopDistance = embedding.distance(between: "computer", and: "laptop")

let neighbors = embedding.neighbors(for: "computer",
                                    maximumCount: 5,
                                    distanceType: .cosine)
```

TRANSFER LEARNING

- Usage of WordEmbeddings to **improve models**
- **Previous knowledge** of the language (pre-trained model)
- **Less training data** required
- Types: **static, dynamic** and **custom**

TRANSFER LEARNING

```
// Creating the model

let params = MLTextClassifier.ModelParameters(algorithm: .maxEnt(revision: 1))

let model  = try MLTextClassifier(trainingData: trainingSet,
                                  textColumn: "text",
                                  labelColumn: "label",
                                  parameters: params)
```

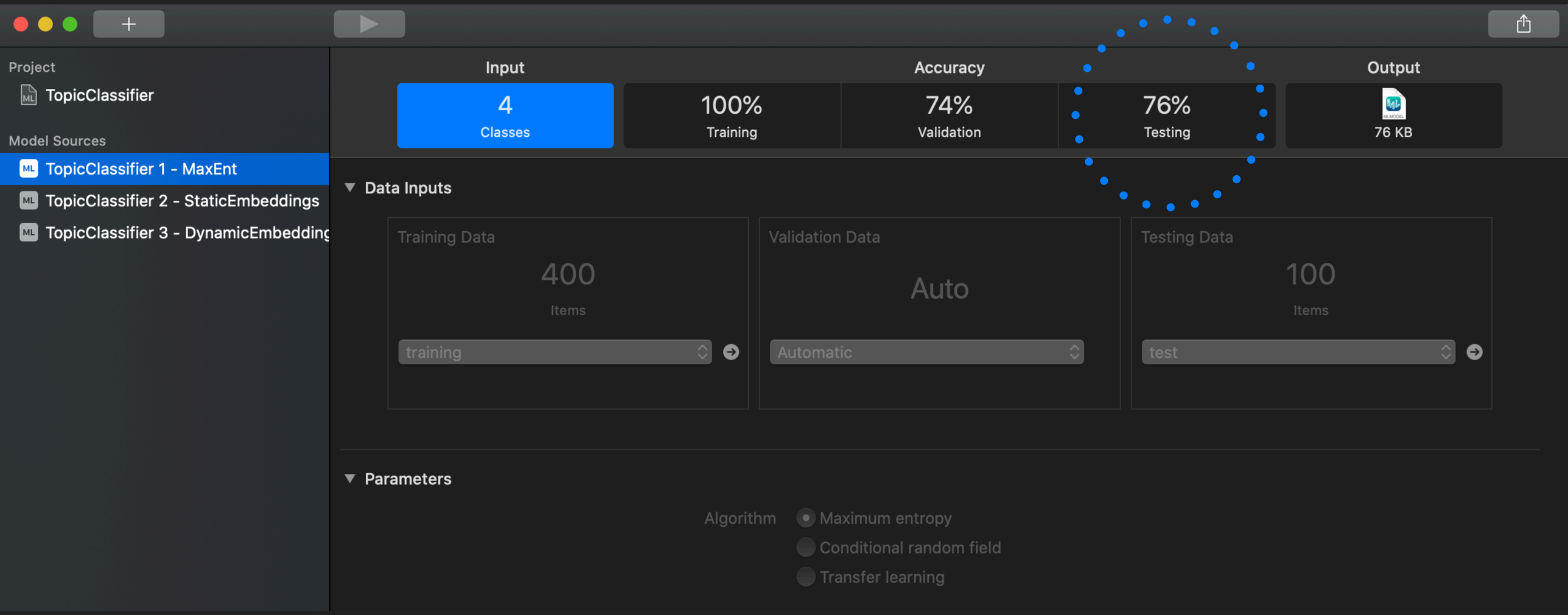
TRANSFER LEARNING

```
// Creating the model

let params = MLTextClassifier.ModelParameters(algorithm: .transferLearning(.staticEmbedding,
                                                                           revision: 1))

let model  = try MLTextClassifier(trainingData: trainingSet,
                                  textColumn: "text",
                                  labelColumn: "label",
                                  parameters: params)
```

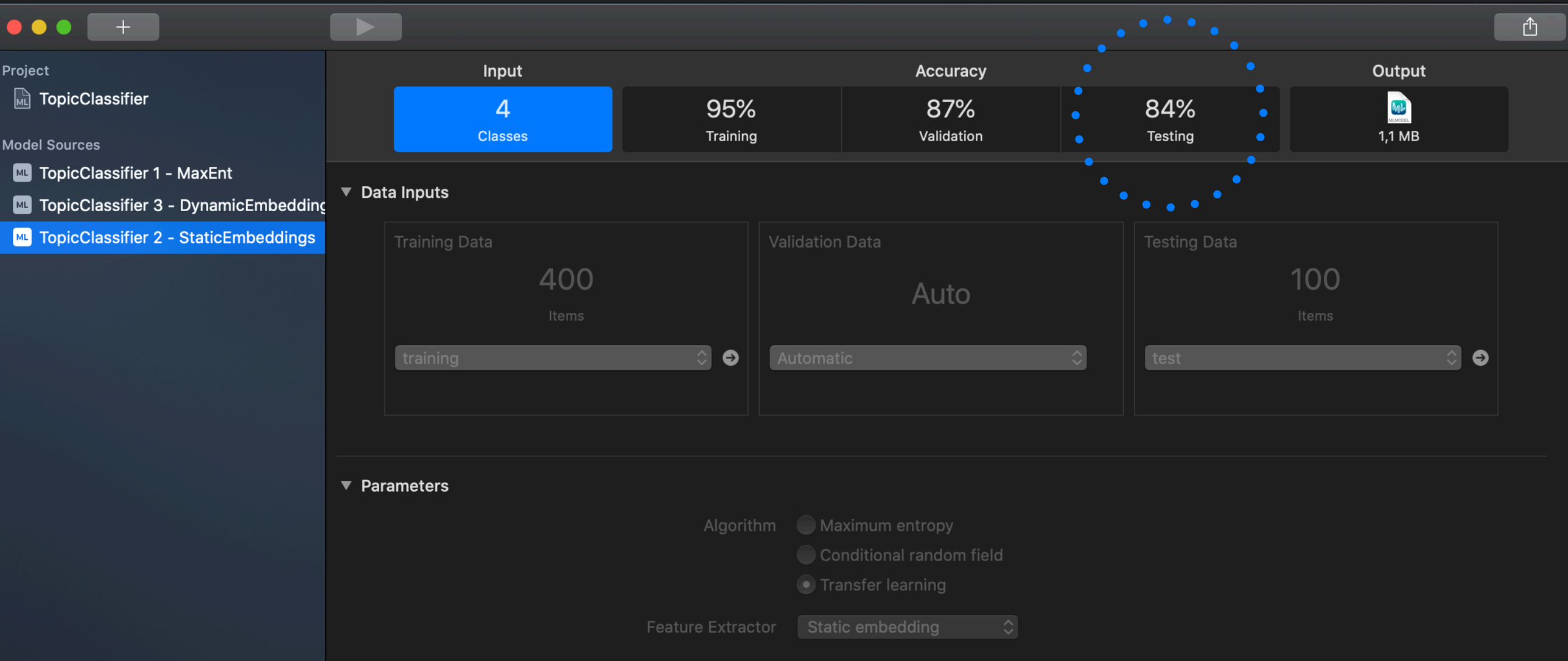
DEMO


TURNIP 2000
INITIATING....WAIT...WAIT A BIT MORE... ALMOST
FUNK SUBROUTINE V2.09
WELCOME TO FUNKTOWN, GROOVER
DEFINE (FUNK PARAMTERS: SUPER CRUNKY
WORKING.....
(CRUNKYness SET TO SETTING: SUPER(SUPER CRUNKY)
C:/LOVE_THE_WAY_YOU_MOVE.MP3 -LOAD TO EXT. DEVICE
WRITTEN SUCCESSFULLY TO EXT. DEVICE
C:/DROP_DA_BEAT.EXE -P(LOVE THE_WAY_YOU_MOVE.MP3)
BEAT DROPPING.....
BEAT SUCCESSFULLY DROPPED. IS IT FRESH? Y/N
Y
TIGHT. INITIATE FLY SOVIET DANCE KICKS TESTING
PHASE? Y/N
Y
INITIATING....PLEASE WAIT.....
.....
SUCCESSFULLY INITIATED.
C:/BASH
BASH 3.20 LX -ART
FUNK DROPS MAD BEATS FLY KICKS
DANCE POPS LOCKS CRUNK
BASH 3.2: CD DANCE

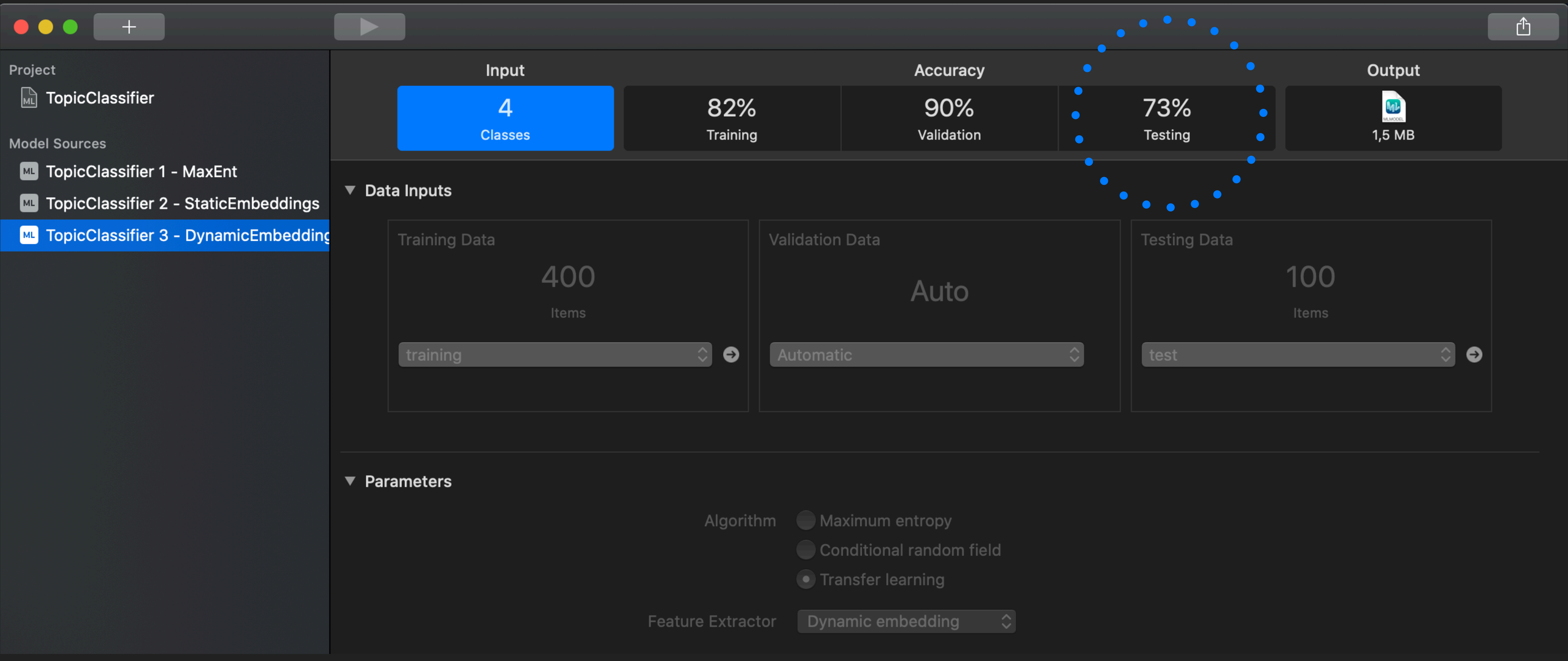
TRANSFER LEARNING



TRANSFER LEARNING



TRANSFER LEARNING



RECAP

Natural Language
Machine Learning

➧ **Text classifier**

- WordEmbeddings
- Transfer learning

Thank you

GITHUB REPO

THE
DEVELOPER'S
CONFERENCE