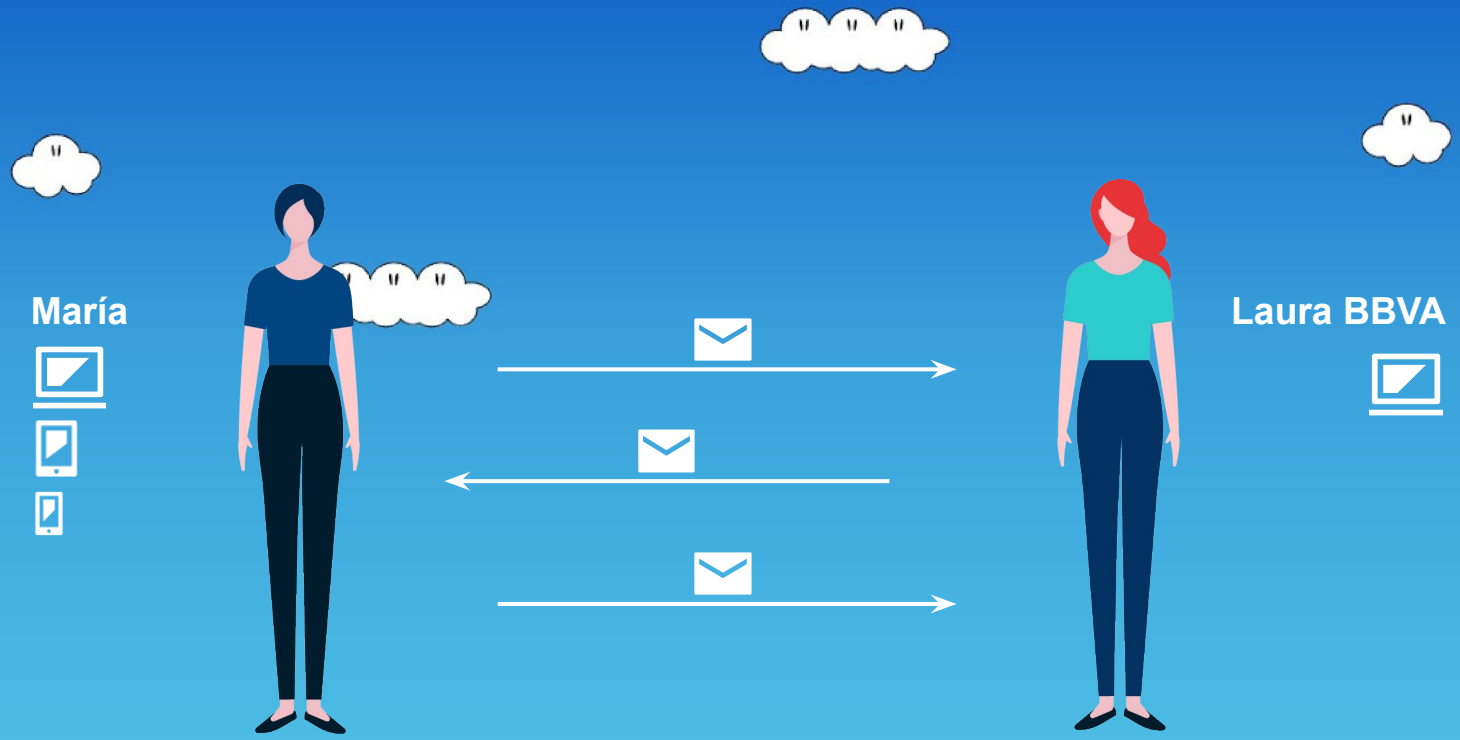
BBVA
Data & Analytics

# NLP adventures in a financial organisation

## From business need to model production

**Clara Higuera, PhD**
Data Scientist
BBVA Data&Analytics




**App BBVA**

Conversations assistant manager client via text messages

**Main goal**

Assist managers with AI to decrease average response time per interactions with client

Natural Language processing

Business need


Model in production

# Who´s playing?

**Ángela Gonzalo**
Theme Leader

**PROGRAMA SMART ASSISTANCE**

**Jesús Martín**
Program Manager

**Beatriz Cano**
Scrum Master

**Laura Díaz**
AI Architect

**Juan Martín Fornell**
Data Specialist

**SCRUMS**

## Smart Conversations

**María Hernández**
Product Owner
Data Scientist

**Pedro Hernández**
Data Scientist

**Javier Asúa**
Data Science Inter

**Cristina Sanz**
Machine Learning Engineer

**Miguel Ángel Gonzá**
Machine Learning Engineer

## Smart Desktop

**César de Pablo**
Product Owner
Data Scientist

**Clara Higuera**
Data Scientist

**Manuel Martín**
Machine Learning Engineer

**Raúl Carvajal**
Machine Learning Engineer

*Chapters*

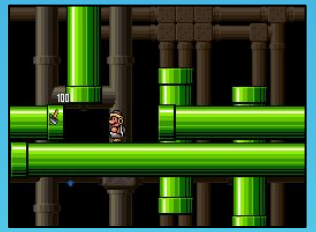
- Business Translators
- Lean Office
- Engineering
- Architecture
- Advanced Analytics
- Data Hub

# A real use case:
## Conversations classifier

Problem definition

What data are available?

Data ingestion & preprocessing



Model training and validation

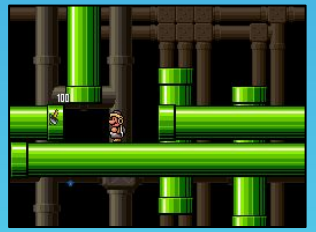
Model into production

# A real use case:
## Conversations classifier

**Problem definition**

**What data are available?**

**Data ingestion & preprocessing**



**Model training and validation**

**Model into production**

# Problem definition

## Conversations classifier

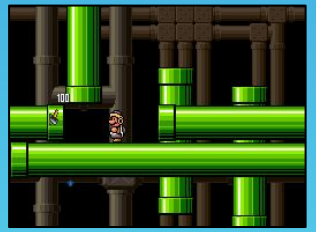
| | | product | operation |
|---|---|---|---|
| Hola, me gustaría cancelar mi tarjeta |  | tarjeta | cancelación |
| Buenas tardes, querría renovar el seguro de casa. Un saludo |  | seguro vivienda | renovación |
| Buenas, te adjunto aquí la copia de mi nuevo dni ya que lo renové hace una semana. ¡gracias! |  | documentación | actualización |

# A real use case:
## Conversations classifier

Problem definition

What data are available?

Data ingestion & preprocessing

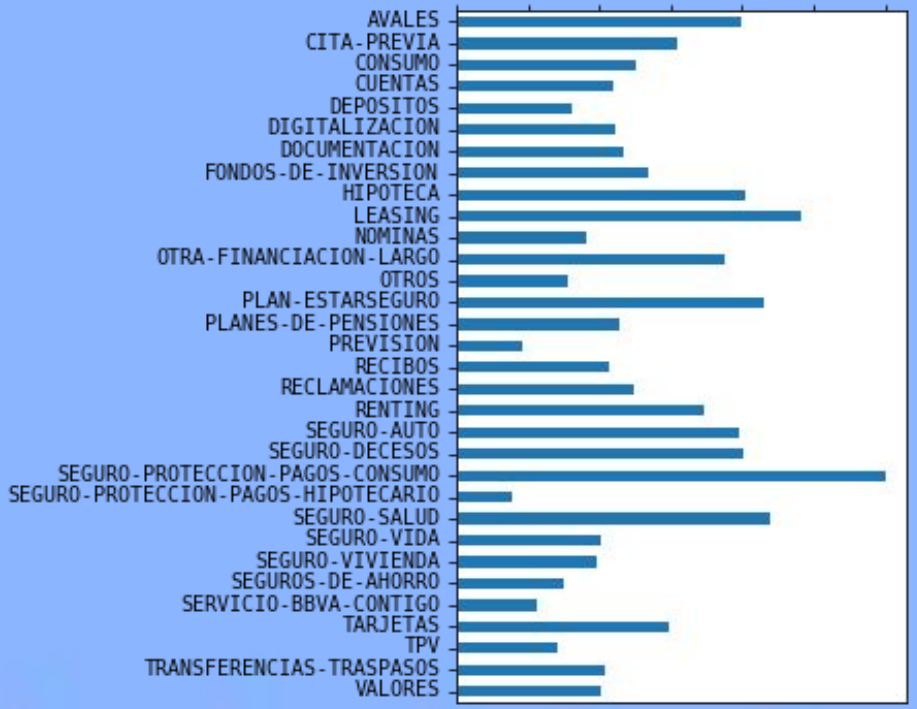


Model training and validation

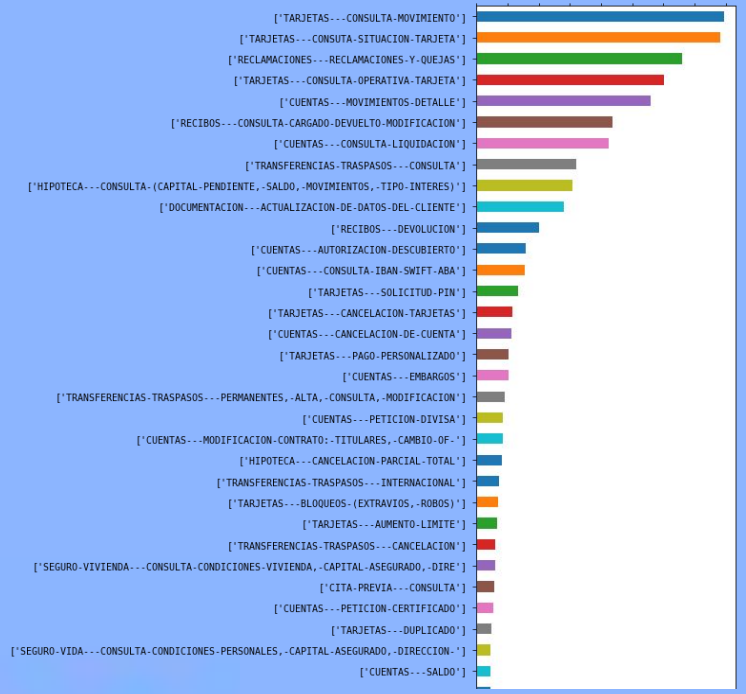
Model into production

# What data do we have?

## 32 products aprox




## 120 operations aprox



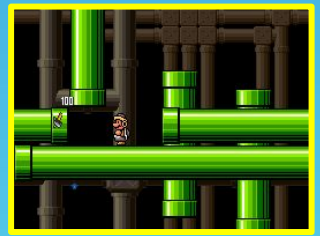

**200k conversations labeled by call center**

# A real use case:
## Conversations classifier

Problem definition

What data are available?

Data ingestion & preprocessing



Model training and validation

Model into production

# Ingestion and preprocessing NLP - ETL pipeline

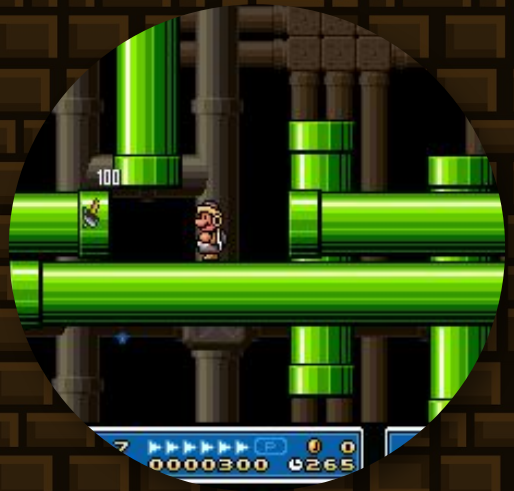
Language identification
Name entity recognition
Anonymization
Text cleaning

Messages

Conversations

Assistant manager info

Conversation labels



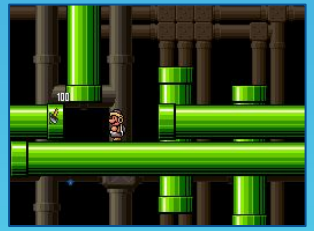
| Conv id | text | product | operation |
|---|---|---|---|
| | | | |
| | | | |
| | | | |

# A real use case:
## Conversations classifier

Problem definition

What data are available?

Data ingestion & preprocessing



Model training and validation

Model into production

# Model training and validation

## How do we train a model with text?
## How do we translate text into a numerical representation?

### Bag of words

```
John likes to watch movies. Mary likes movies too.
. Mary also likes to watch football games.
```

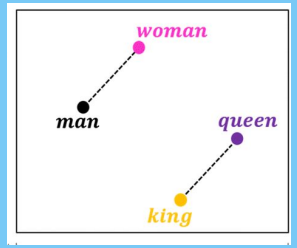
```
{"John":1,"likes":3,"to":2,"watch":2,"movies":2,"Mary":2,"too":1,"also":1,"football":1,"games":1}
```

**Tf-idf - Term frequency – Inverse document frequency**

$$w_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right)$$

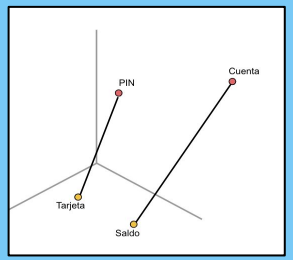
$tf_{i,j}$ = number of occurrences of $i$ in $j$
$df_i$ = number of documents containing $i$
$N$ = total number of documents

### Word embeddings




Models that map words or phrases from the vocabulary to vectors of real numbers.

Takes context into account




**Examples of models**

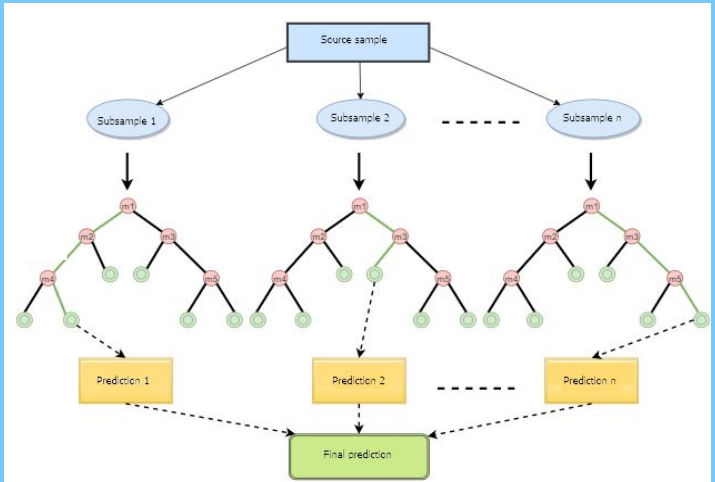
Word2vec

GloVe

https://vishwamittra.github.io/index.html

# Model training and validation

## Random Forest



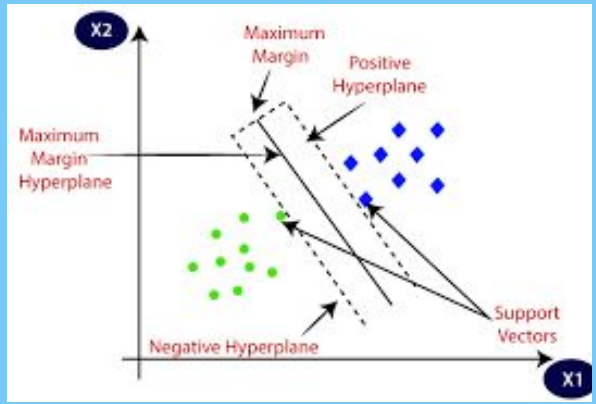

## Naive Bayes

$$P(A|B) = \frac{P(B|A)\ P(A)}{P(B)}$$

using Bayesian probability terminology, the above equation can be written as

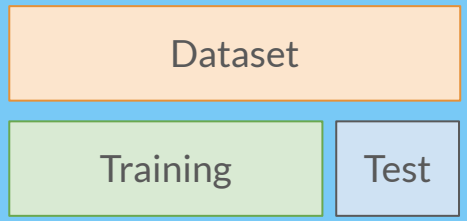
$$\text{Posterior} = \frac{\text{prior x likelihood}}{\text{evidence}}$$

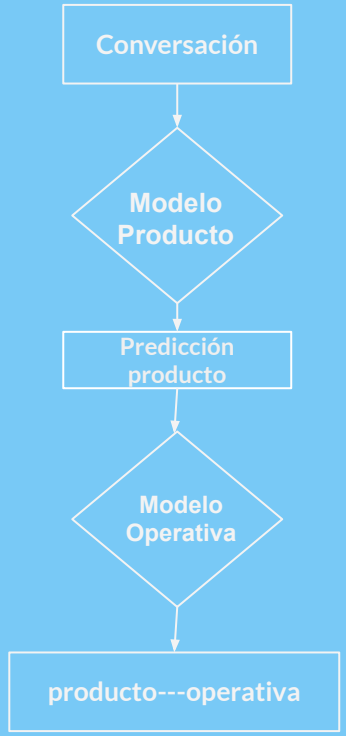
## Support vector machine



# Model training and validation

## Classic validation

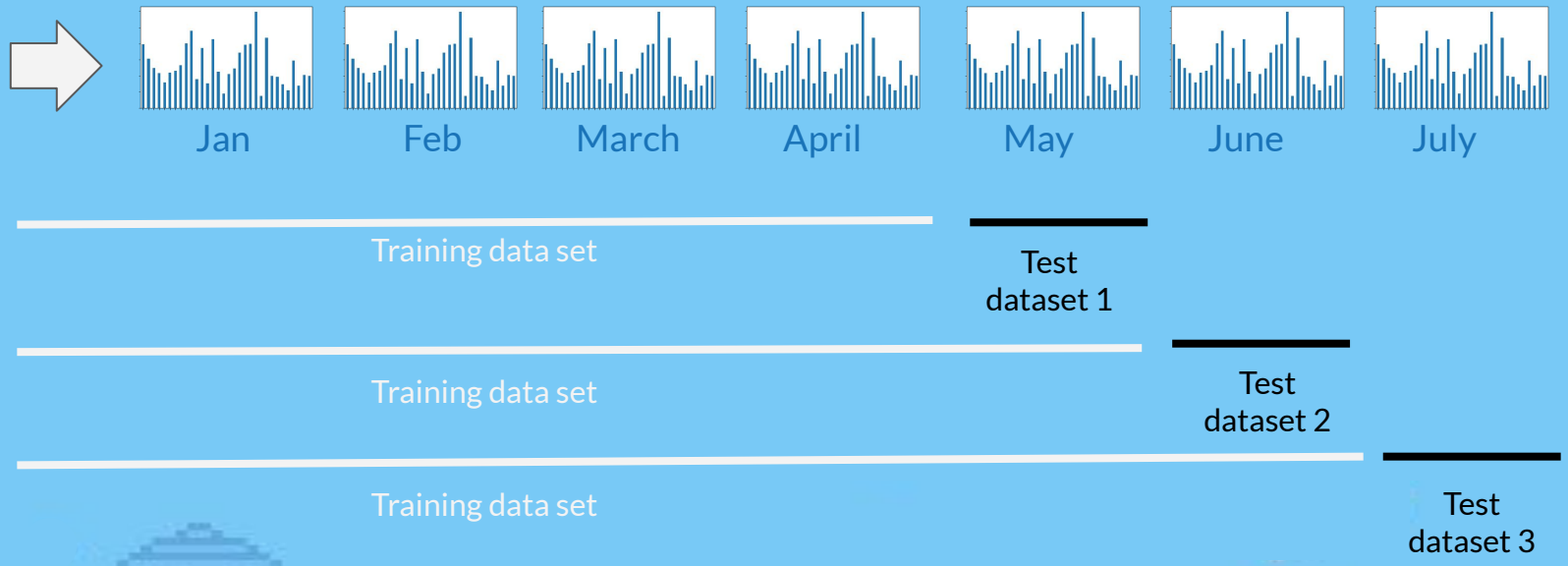



| | precision | recall | f1-score |
|---|---|---|---|
| avales | 1.00 | 0.33 | 0.50 |
| cita-previa | 0.92 | 0.79 | 0.85 |
| consumo | 0.76 | 0.55 | 0.64 |
| cuentas | 0.71 | 0.80 | 0.75 |
| depositos | 0.65 | 0.58 | 0.61 |
| digitalizacion | 0.73 | 0.46 | 0.57 |
| documentacion | 0.73 | 0.81 | 0.77 |
| fondos-de-inversion | 0.82 | 0.70 | 0.75 |
| hipoteca | 0.83 | 0.80 | 0.82 |
| leasing | 0.00 | 0.00 | 0.00 |
| nominas | 0.33 | 0.16 | 0.22 |
| otros | 0.39 | 0.45 | 0.42 |
| plan-estarseguro | 0.79 | 0.30 | 0.43 |
| planes-de-pensiones | 0.80 | 0.70 | 0.75 |
| prevision | 0.88 | 0.37 | 0.52 |
| recibos | 0.92 | 0.88 | 0.90 |
| reclamaciones | 0.57 | 0.38 | 0.46 |
| renting | 0.00 | 0.00 | 0.00 |
| seguro-auto | 0.64 | 0.56 | 0.60 |
| seguro-decesos | 0.00 | 0.00 | 0.00 |
| seguro-proteccion-pagos-consumo | 0.00 | 0.00 | 0.00 |
| seguro-proteccion-pagos-hipotecario | 0.00 | 0.00 | 0.00 |
| seguro-salud | 0.50 | 0.18 | 0.26 |
| seguro-vida | 0.68 | 0.45 | 0.54 |
| seguro-vivienda | 0.70 | 0.51 | 0.59 |
| seguros-de-ahorro | 0.00 | 0.00 | 0.00 |
| servicio-bbva-contigo | 0.71 | 0.25 | 0.37 |
| tarjetas | 0.92 | 0.96 | 0.94 |
| tpv | 0.92 | 0.59 | 0.72 |
| transferencias-traspasos | 0.88 | 0.88 | 0.88 |
| valores | 0.94 | 0.84 | 0.88 |
| avg / total | 0.83 | 0.84 | 0.83 |

# Model training and validation

## Temporal validation



# Model training and validation

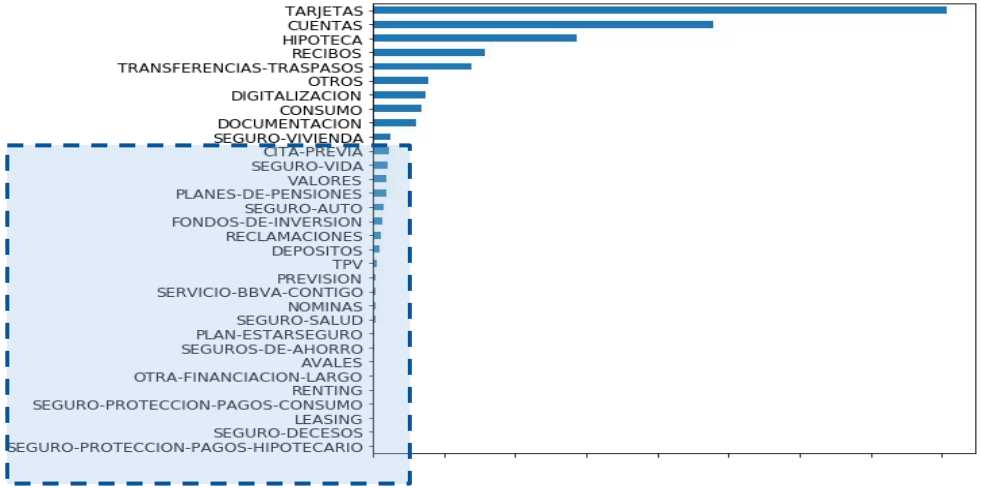
## Manual validation

**Contact center**

Cuentas
Tarjetas
Recibos
Transferencias

**Financial advisor**

Hipotecas
Fondos de Inversión
Seguro vivienda
Fondos de inversión



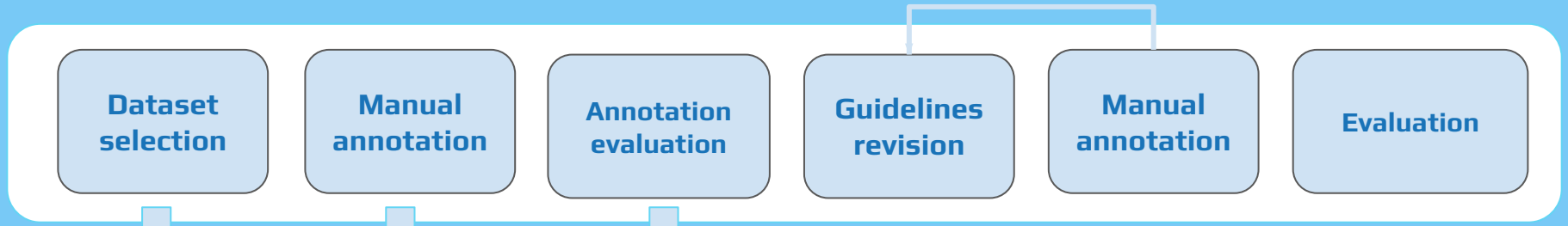

Data labeled by contact center but model will be used by financial advisors!

Ho do we estimate the performance of the model in production?

Human bias: sometimes humans do not label as thoroughly as needed for training models

# Model training and validation

## Manual validation




360 conversations not labeled

120 conversations

1 conv / 12 annotators

$$\kappa \equiv \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e},$$

| Kappa value | Strength of agreement | Distribution of kappa values | |
|---|---|---|---|
| | | Intra-rater agreement (%) | Inter-rater agreement (%) |
| <0.01 | Poor | 1 (4.3) | 0 (0.0) |
| 0.01–0.20 | Slight | 0 (0.0) | 0 (0.0) |
| 0.21–0.40 | Fair | 0 (0.0) | 2 (9.5) |
| 0.41–0.60 | Moderate | 3 (13.1) | 3 (14.3) |
| 0.61–0.80 | Substantial | 4 (17.4) | 4 (19.0) |
| 0.81–1.00 | Almost perfect | 15 (65.2) | 12 (57.1) |
| | Total | 23 (100.0)[a] | 21 (100.0)[b] |

[a] One item (no. 20. in Table 1) could not be calculated
[b] Three items (nos. 6, 8, and 20 in Table 1) could not be calculated

# Model training and validation

## Manual validation

**Dataset selection** → 360 conversations not labeled

**Manual annotation** → 120 conversations

1 conv / 12 annotators

**Annotation evaluation**

$$\kappa \equiv \frac{p_o - p_e}{1 - p_e} = 1 - \frac{1 - p_o}{1 - p_e},$$

| | Cohen-Kappa |
|---|---|
| Angela-Juan | 0.760372 |
| Jesus-Angela | 0.678398 |
| Cris-Jesus | 0.676972 |
| Jesus-Pedro | 0.675780 |
| Jorge-Cris | 0.667735 |
| Jorge-Pedro | 0.653369 |
| Cris-Juan | 0.644666 |
| Maria-Cris | 0.610090 |
| Cris-Pedro | 0.589880 |

**Average: 0.45**

**Guidelines revision** → **Manual annotation** → **Evaluation**

# Model training and validation

## Manual validation

**Dataset selection**

360 conversations not labeled

**Manual annotation**

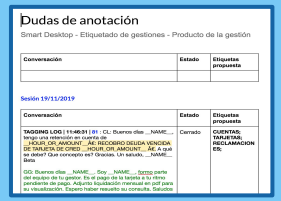
120 conversations

1 conv / 12 annotators

**Annotation evaluation**

| | Cohen-Kappa |
|---|---|
| Angela-Juan | 0.760372 |
| Jesus-Angela | 0.678398 |
| Cris-Jesus | 0.676972 |
| Jesus-Pedro | 0.675780 |
| Jorge-Cris | 0.667735 |
| Jorge-Pedro | 0.653369 |
| Cris-Juan | 0.644666 |
| Maria-Cris | 0.610090 |
| Cris-Pedro | 0.589880 |
| Cris-Angela | 0.586231 |
| Angela-Pedro | 0.575460 |
| Jesus-Juan | 0.542189 |
| Pedro-Laura | 0.541786 |

**Guidelines revision**




**Manual annotation**

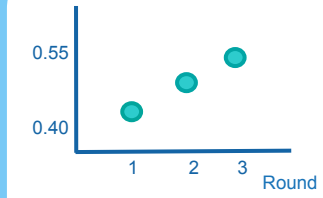
2400 conversations

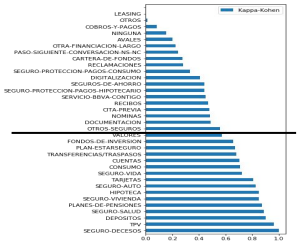
200 conv x 12 annotators

16 prod

**Cohen Kappa >=0.6**

**Evaluation**



# Model training and validation

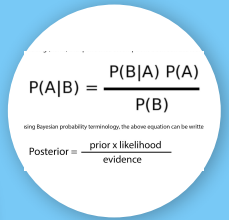
## Three step validation

| Modelo multiclass | Dataset | Precision | Recall |
|---|---|---|---|
| RF - varios clasificadores | CC - validation | 0.90 | 0.66 |
| RF - 1 clasificador | CC-validation | **0.80** | **0.81** |
| RF-varios clasificadores | Gestores | 0.68 | 0.25 |
| RF-1 clasificador | Gestores | **0.66** | **0.64** |
| SVM - 1 Clasificador | CC-validation | 0,83 | 0,84 |
| SVM - 1 Clasificador | Gestores | 0.71 | 0.71 |

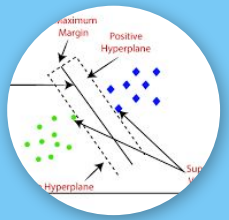
Gestores: Manually annotated dataset

Random Forest

Support vector machine




Naive Bayes



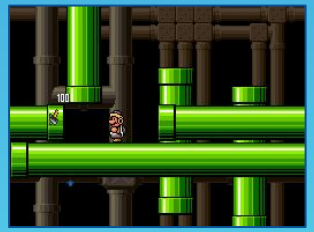

2 1 3

# A real use case:
## Conversations classifier

Problem definition

What data are available?

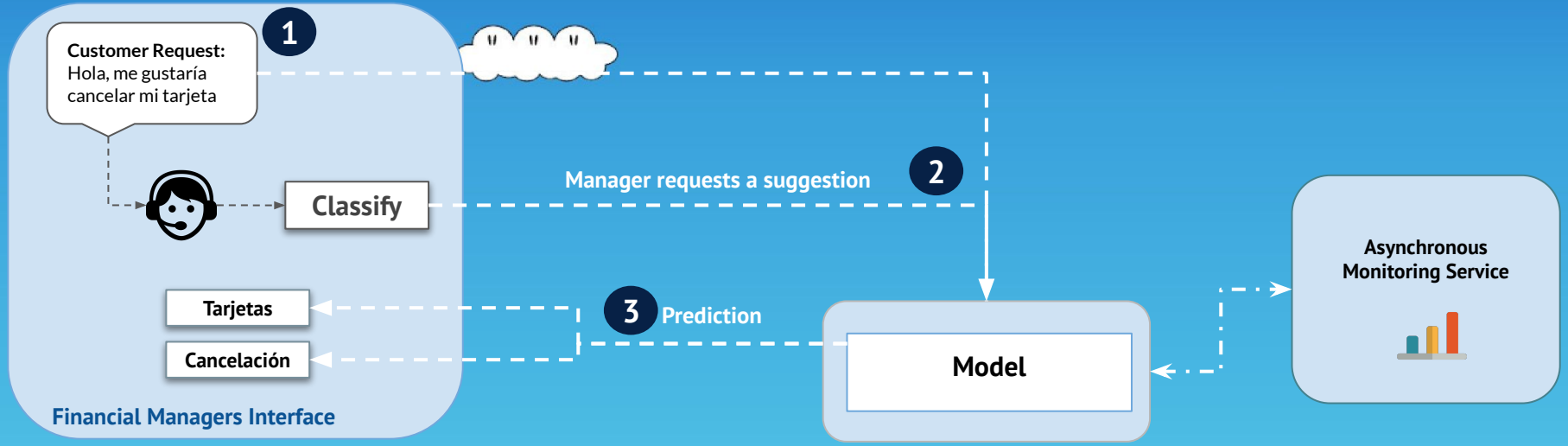
Data ingestion & preprocessing



Model training and validation

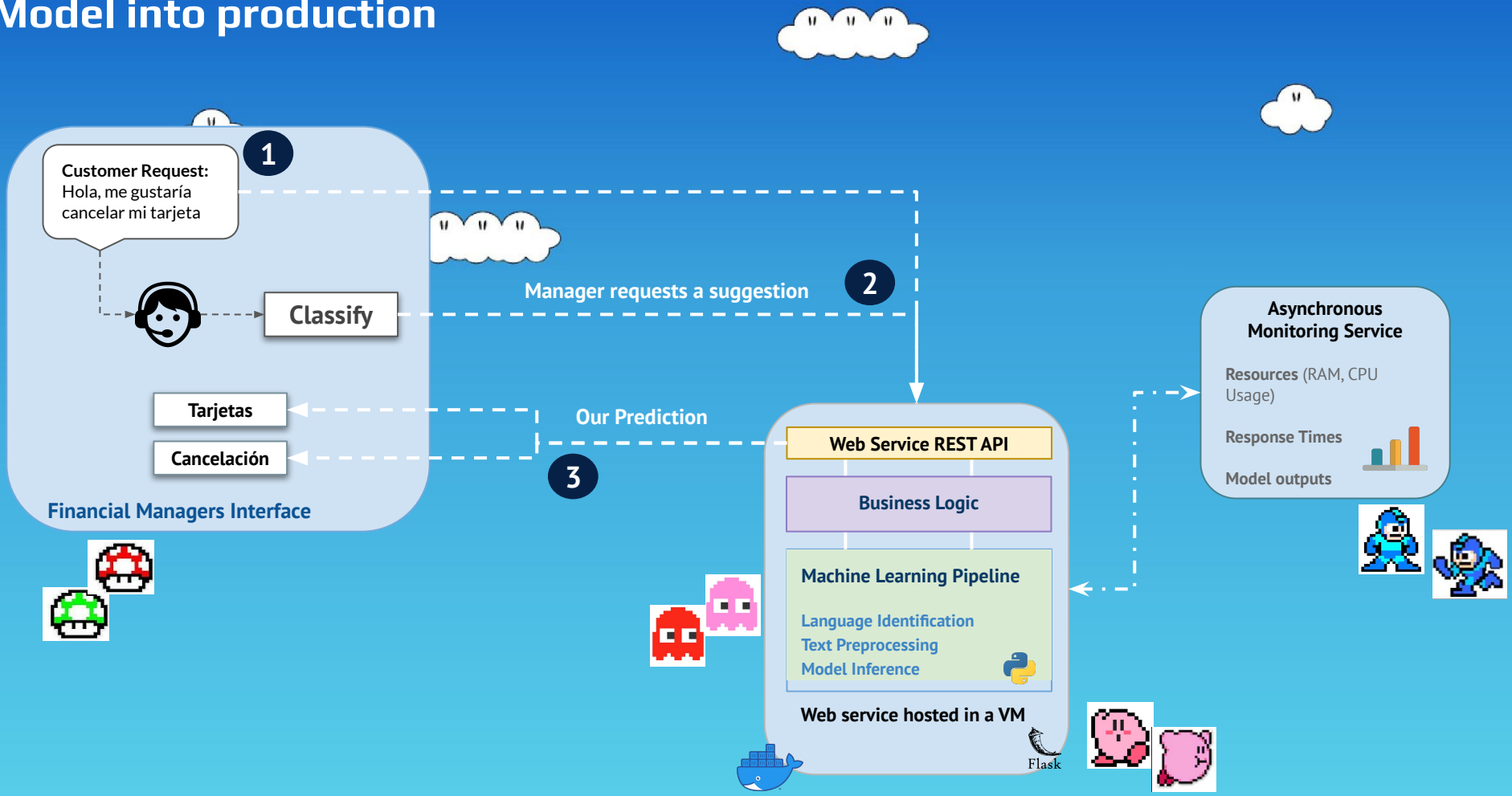
Model into production

# Model into production




maintenance and improvement

# Model into production



Problem definition

What data are available?

Data ingestion & preprocessing

Model training and validation

Model into production


That´s all folks!