



Universitat Oberta  
de Catalunya

# Análisis predictivo del comportamiento de los clientes en sus interacciones con la empresa

Diego Contreras Jiménez

Barcelona, 15 de Junio de 2019

# 01 INTRODUCCIÓN

01

INTRODUCCIÓN

02

ESTADO DEL  
ARTE

03

DISEÑO Y  
DESARROLLO

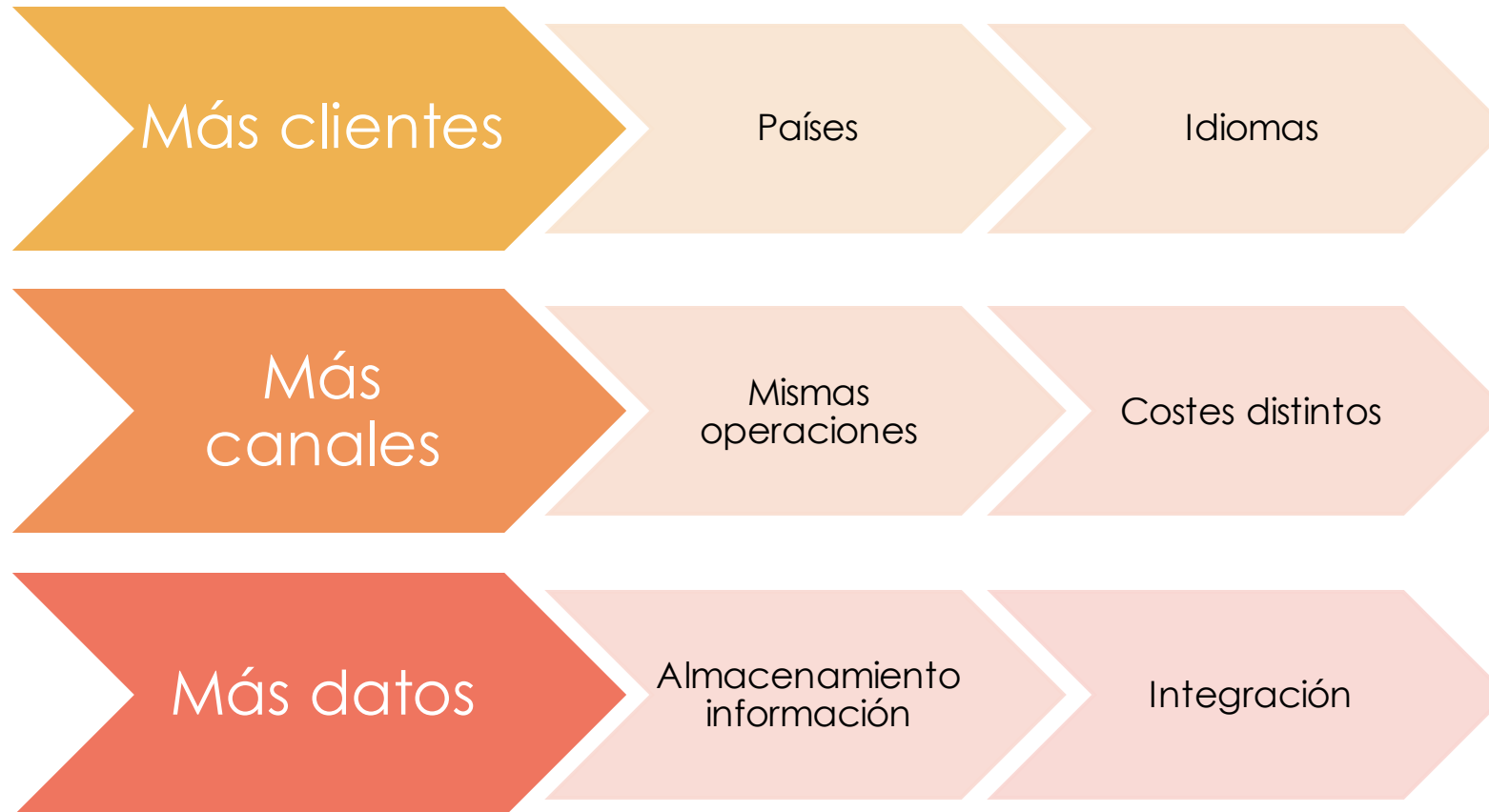
04

EXPERIMENTOS  
Y RESULTADOS

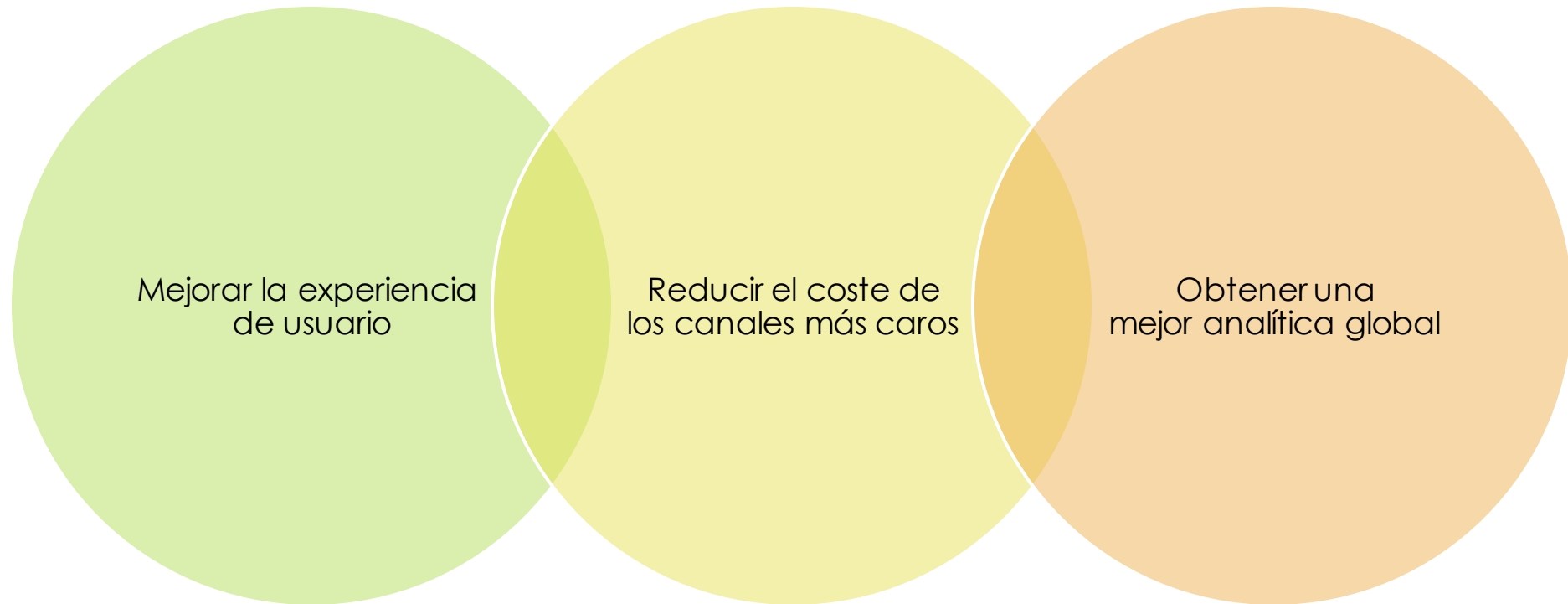
05

CONCLUSIONES

# 01 INTRODUCCIÓN



# 01 INTRODUCCIÓN



# 02 ESTADO DEL ARTE

01

INTRODUCCIÓN

02

ESTADO DEL  
ARTE

03

DISEÑO Y  
DESARROLLO

04

EXPERIMENTOS  
Y RESULTADOS

05

CONCLUSIONES

# 02 ESTADO DEL ARTE

## Big Data

- Conjunto de estrategias, tecnologías y sistemas para el almacenamiento, procesamiento, análisis y visualización de conjuntos de datos complejos

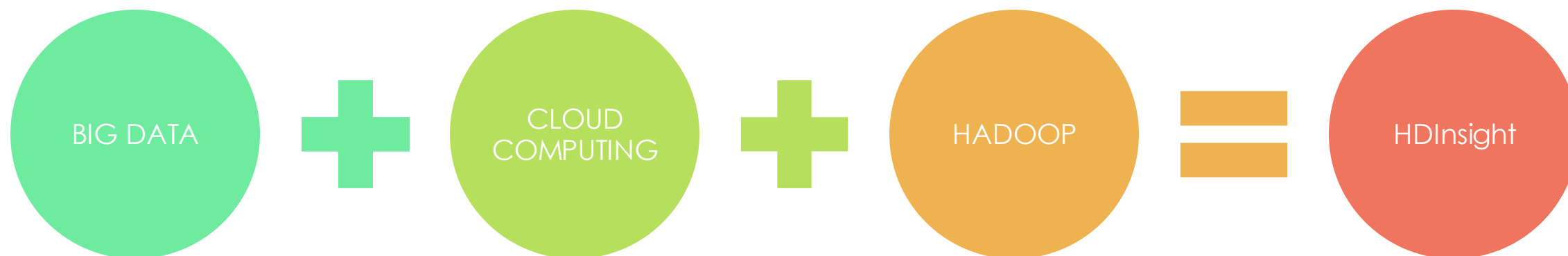
## Cloud Computing

- Infrastructure as a service (IaaS)
- Platform as a service (PaaS)
- Software as a service (SaaS)

## Hadoop

- Almacenamiento HDFS
- Apache Spark
- Machine Learning

## 02 ESTADO DEL ARTE



# 03 DISEÑO Y DESARROLLO

01

INTRODUCCIÓN

02

ESTADO DEL  
ARTE

03

DISEÑO Y  
DESARROLLO

04

EXPERIMENTOS  
Y RESULTADOS

05

CONCLUSIONES



# 03 DISEÑO Y DESARROLLO



# 03 DISEÑO Y DESARROLLO

Logistic  
Regression

Decision Tree  
Classifier

Random  
Forest  
Classifier

Gradient-  
boosted Tree  
Classifier

Linear Support  
Vector  
Machine

# 04 EXPERIMENTOS Y RESULTADOS

01

INTRODUCCIÓN

02

ESTADO DEL  
ARTE

03

DISEÑO Y  
DESARROLLO

04

EXPERIMENTOS  
Y RESULTADOS

05

CONCLUSIONES

# 04 EXPERIMENTOS Y RESULTADOS

## Casos de uso

- Predicción de facturación electrónica
- Predicción del canal por el que se ha realizado una reclamación

# 04 EXPERIMENTOS Y RESULTADOS

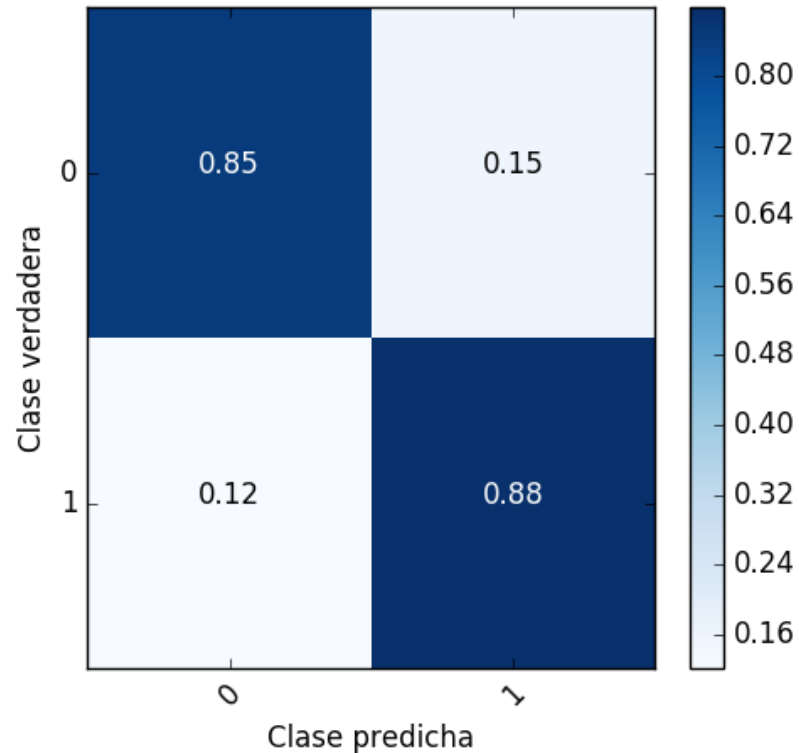
## Predicción de facturación electrónica

- **Linear Support Vector Machine**
  - Exactitud (ACC): 0.86
  - Error de clasificación(ERR): 0.14
  - Tasa de verdaderos positivos o sensibilidad (TPR | REC): 0.88
  - Tasa de verdaderos negativos (FPR): 0.15
  - Precisión (PRE): 0.85
  - Especificidad(SPE): 0.85
  - F1: 0.86
  - Área bajo la curva ROC: 0.93

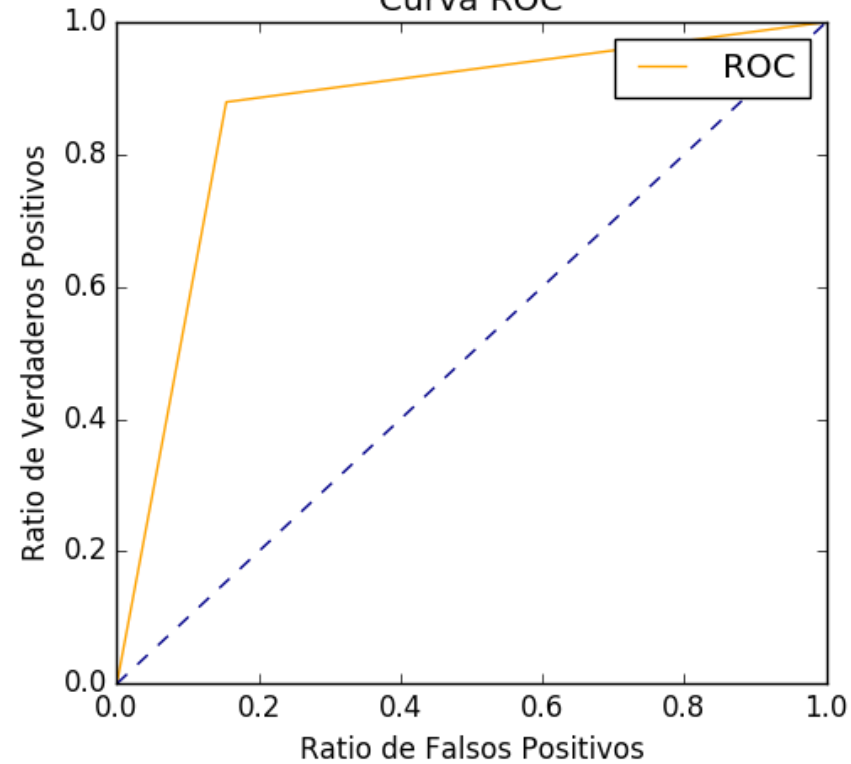
# 04 EXPERIMENTOS Y RESULTADOS

Predicción de facturación electrónica

Matriz de confusión normalizada



Curva ROC



# 04 EXPERIMENTOS Y RESULTADOS

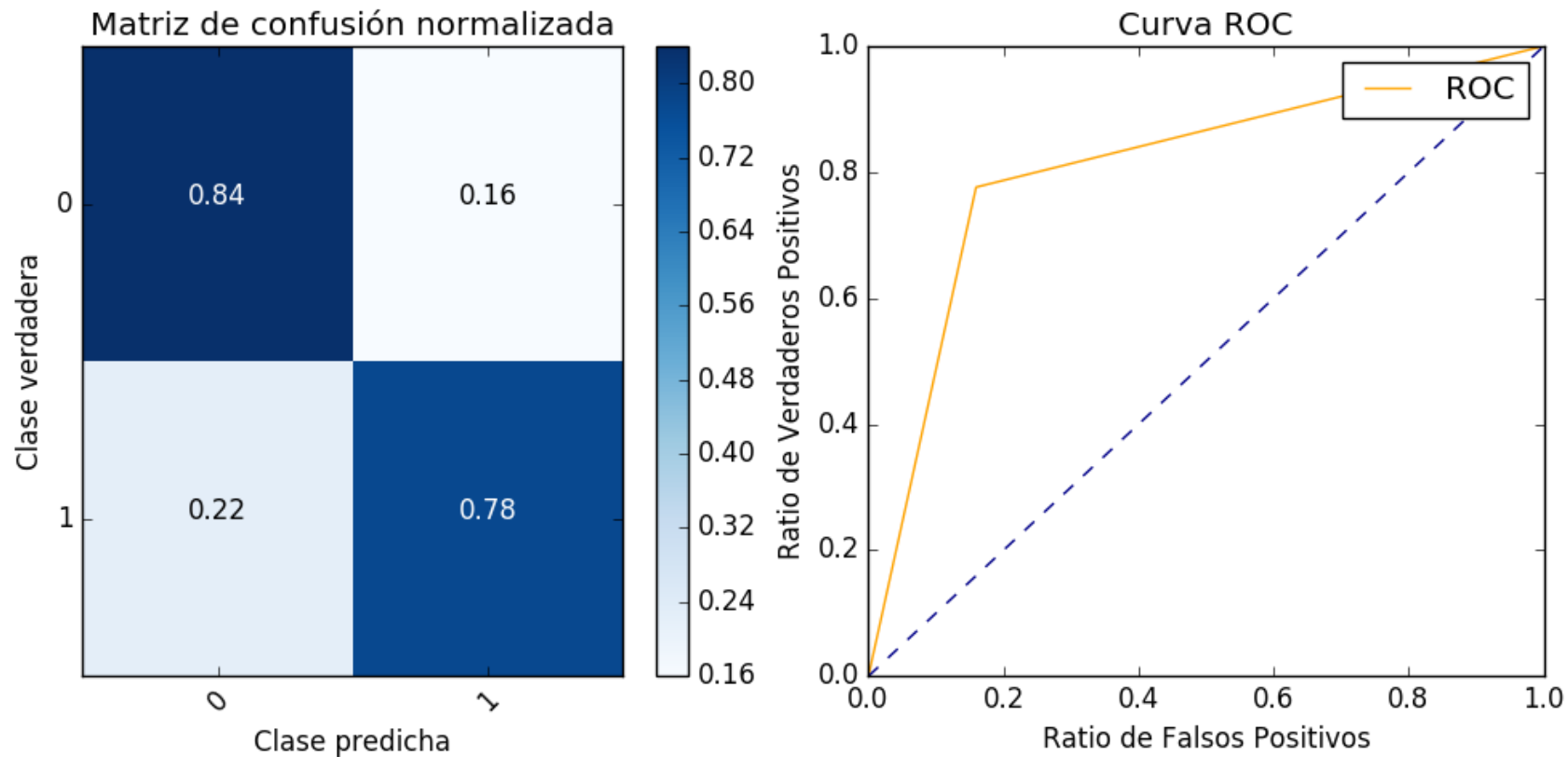
Predicción del canal por el que se ha realizado una reclamación

- **Gradient-boosted Tree Classifier**

- Exactitud (ACC): 0.81
- Error de clasificación (ERR): 0.19
- Tasa de verdaderos positivos o sensibilidad (TPR | REC): 0.78
- Tasa de verdaderos negativos (FPR): 0.16
- Precisión(PRE): 0.83
- Especificidad(SPE): 0.84
- F1: 0.80
- Área bajo la curva ROC: 0.89

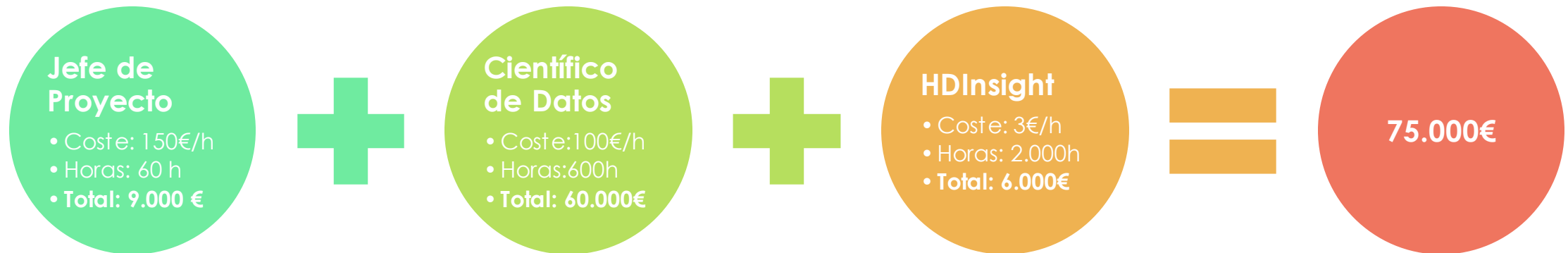
# 04 EXPERIMENTOS Y RESULTADOS

Predicción del canal por el que se ha realizado una reclamación





# 04 EXPERIMENTOS Y RESULTADOS



# 05 CONCLUSIONES

01

INTRODUCCIÓN

02

ESTADO DEL  
ARTE

03

DISEÑO Y  
DESARROLLO

04

EXPERIMENTOS  
Y RESULTADOS

05

CONCLUSIONES

# 05 CONCLUSIONES

## Ciclo completo de un proyecto analítico de Big Data

- Dos casos de uso
- Datos reales de la empresa

## Máster Ciencia de Datos

- Aplicado conceptos vistos en otras asignaturas
- Ampliación de conocimientos

## Otras habilidades

- Python
- Machine Learning
- Agile
- LaTeX

# 05 CONCLUSIONES

## Líneas de trabajo futuras

- Otros modelos predictivos para otras clases
- Utilizar otros modelos de Machine Learning
- Desplegar por completo el proceso en el entorno Azure HDInsight

# GRACIAS

Diego Contreras Jiménez  
dieconji@uoc.edu