Deutsche Post DHL Group

RESIDENTIAL IDENTIFICATION

(Big Data)

Andre Wittfoth
2017 Parcel & Postal GIS Summit
Helsinki 5th of April 2017



Agenda

1 **Purpose of project**
Overview of business (OPS) requirements

2 **Set up of possible solution**
Overview of system set up, data sources and results POC

3 **Questions**
Questions & remarks

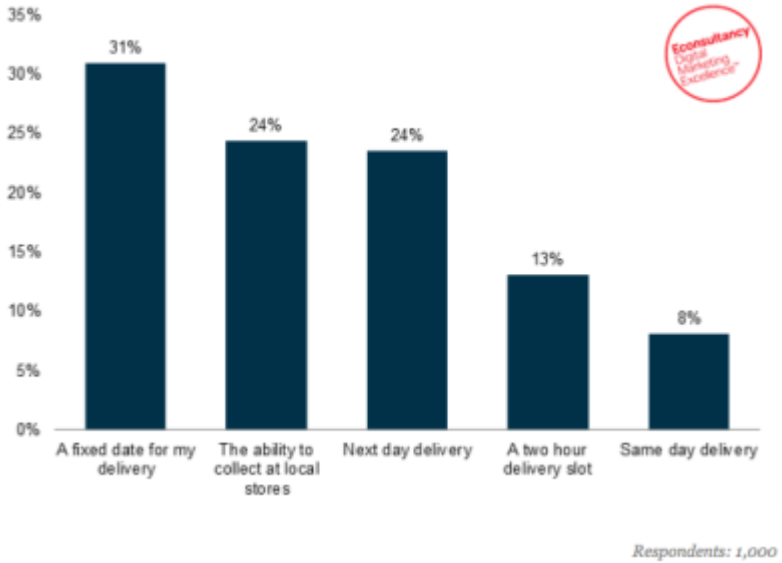
Deutsche Post DHL
Group

PURPOSE OF PROJECT

BUSINESS REQUIREMENTS



Purpose of Residential Identification















Purpose of Residential Identification

Business requirements

- Solution to identify shipments with possible residential delivery
- Solution which can operate with current data quality
- Decrease of manual (double) handling of shipments
- Information on which areas have opportunity of residential delivery during business hours
- Ability to fulfill customer needs regarding delivery time windows

Expected benefits

- Increase of successful 1st attempt deliveries
- Decrease of additional costs for handling failed 1st attempt deliveries
- Increase customer satisfaction



Deutsche Post DHL Group



SET UP OF POSSIBLE SOLUTION

OVERVIEW OF SYSTEM SET UP, DATA SOURCES AND RESULTS POC

Proof of Concept – Data sources

In scope



- Ireland as destination country
- All origin countries sending to Ireland

- Usage of NH* + CA* shipments
- Use 5 years history
- Not existing postal codes
- Already existing manual process for identification
- "If I can make it there I'll make it anywhere"

*NH: Not Home CA: Closed on Arrival

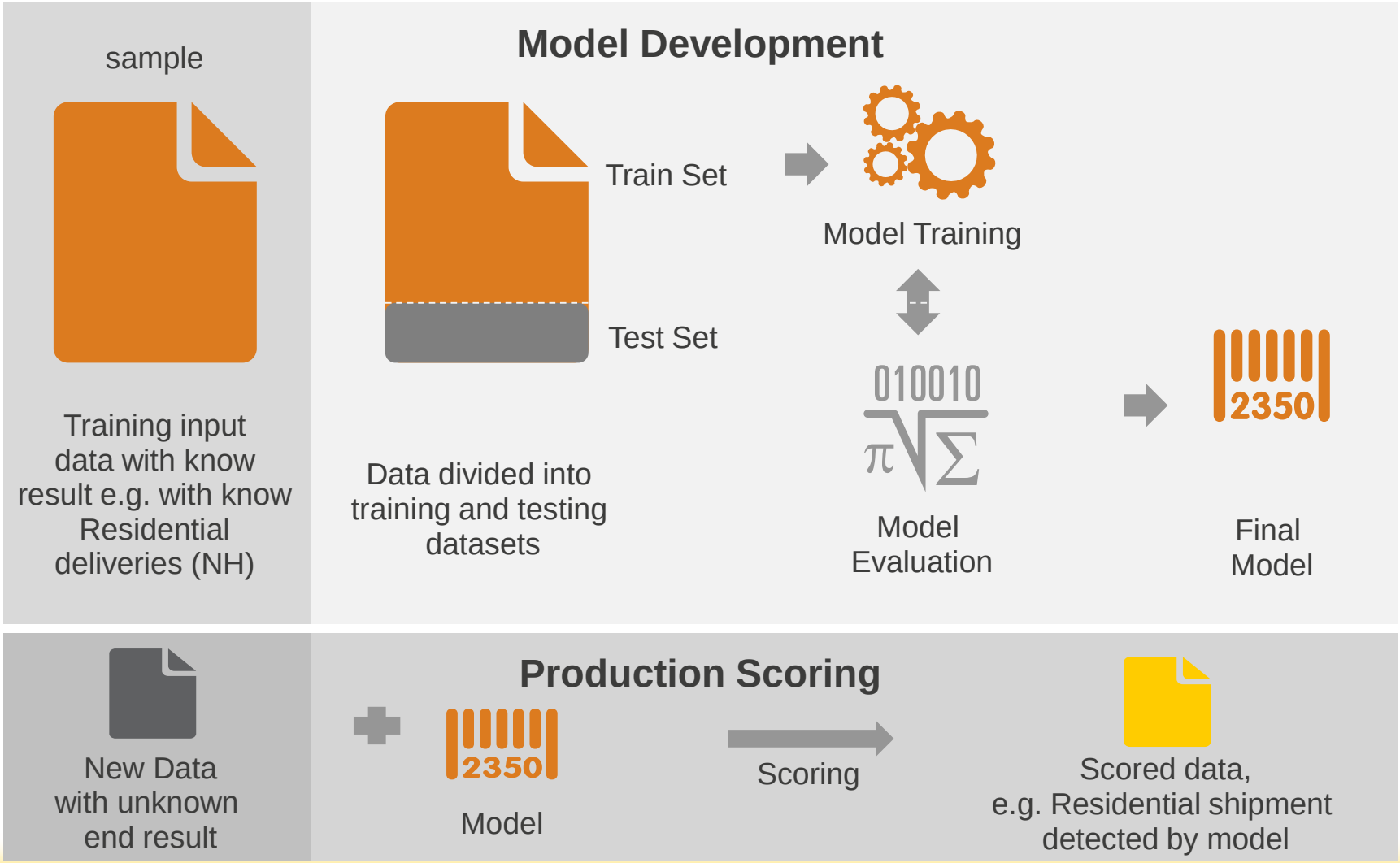
Internal data	**OPMS** (900k records) ▪ Shipment details ▪ 485 attributes
Internal data	**CMF** (500k records) ▪ Customer name ▪ Count company* ident
External data	**Names & surnames** (400k) ▪ Names ▪ Surnames
External data	**Google place APIs** ▪ Name, Street, Town ▪ Opening hours

*company is used for DHL customers



Can we predict (with sufficient accuracy) that the receiver of a shipment is a resident?

Proof of Concept – System Setup



Proof of Concept – Model Training



Score constructed from typically 50-200 particular probabilities combined to final score (text tokens + numeric values + interactions = features)

5.414.213.280 features taken into account







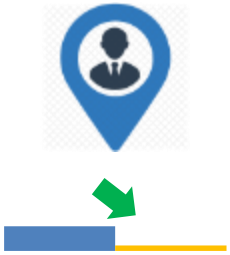
standardized coefficient magnitudes

Name	Importance
Maschine Learning	**67%**
Company identificator	17%
Surnames	5%
Firstnames	4%
Calc_weight	3%
Calc_Piece_No	2%
Diff_name_comp	1%
Google	0.7%
Calc_vol_weight	0.3%

Proof of Concept – Results



correct Residential



wrong classified



Machine Learning
+20% correct Identification
- 80% wrong Classification



Manuall Process
+25% Inbound identified as Residential (6% more than auto)



We successfully built a prediction model that is able to detect correctly 96.4% of all Residential Deliveries.

Deutsche Post DHL
Group

POSSIBLE USE OF GIS INFORMATION

INTERNAL POSTAL CODE SYSTEM WITH DOMINANT DELIVERY TYPE





INTERNAL POSTAL CODE SYSTEM
WITH DOMINANT DELIVERY TYPE

Deutsche Post DHL
Group

POSSIBLE USE OF GIS INFORMATION

BEST DELIVERY TIME (WITHIN BIZ-HOUR, EVENING-HOUR, HOUR OF THE DAY)





BEST DELIVERY TIME (WITHIN BIZ-HOUR, EVENING-HOUR, HOUR OF THE DAY)

Deutsche Post DHL
Group

QUESTIONS & REMARKS



Deutsche Post DHL
Group

THANK YOU

