integris
ICT & Cognitive Solutions

INTEGRIS

THE COMPANY
THE PEOPLE
THE SOLUTIONS

Nicola Procopio

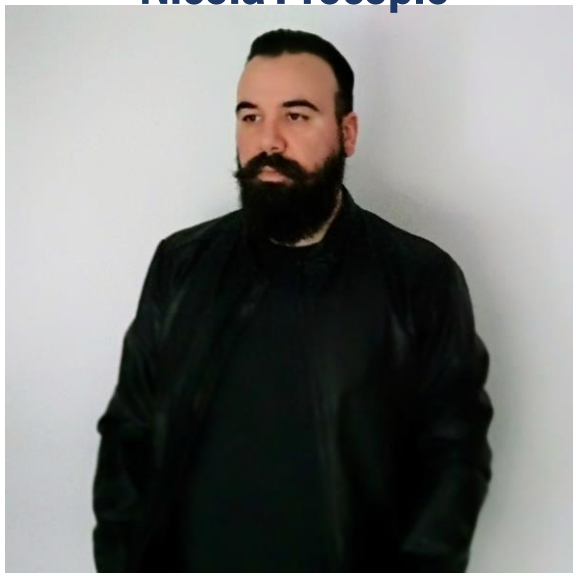
MAY 28th 2019

integris

ICT & COGNITIVE SOLUTIONS

Who I am

Nicola Procopio



Data & Machine Learning Scientist @ integris

My CV

Communities

Tech Interest

Roma
Via Giovanni Squarcina, 3 - 00143 Roma
Tel: +39 06 5032715

Rende
Via Pedro Alvares Cabral, 6 - 87036 Rende (CS)
Tel: +39 0984 395257

Pisa
Via Forti Umberto, 1 - 56121 Pisa (PI)
Tel: +39 050 9655012

Milano
Corso Sempione, 62 - 20149 Milano (MI)
Tel: +39 02 316282

integris

ICT & COGNITIVE SOLUTIONS

The Company

Our Offices

ROMA

PISA

RENDE

MILANO

Our Customers

AXA, Unipol Gruppo, Posteitaliane, CREDEM, TIM, Olivetti, enel, Terna, ADICONSUM, NISSAN, INVITALIA, Rai, Q8, Trenitalia, Aiscat, SEA ambiente, CROSSROAD, everis, IBM, Capgemini, EY, Deloitte., Atos, NTT DATA, indra, Almaviva

Roma
Via Giovanni Squarcina, 3 - 00143 Roma
Tel: +39 06 5032715

Rende
Via Pedro Alvares Cabral, 6 - 87036 Rende (CS)
Tel: +39 0984 395257

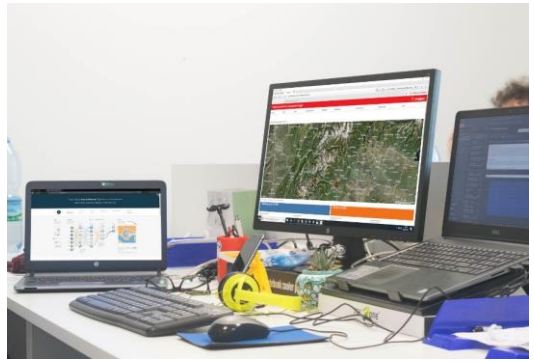
Pisa
Via Forti Umberto, 1 - 56121 Pisa (PI)
Tel: +39 050 9655012

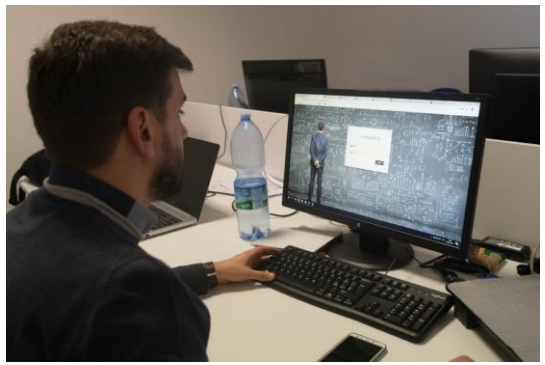
Milano
Corso Sempione, 62 - 20149 Milano (MI)
Tel: +39 02 316282

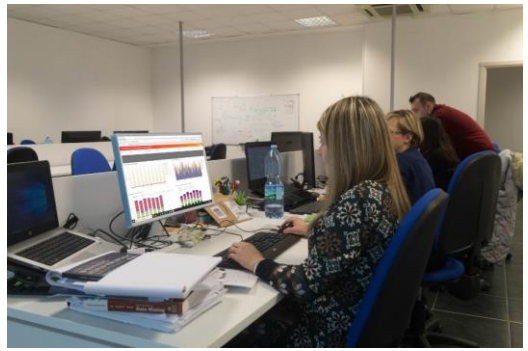
integris

ICT & COGNITIVE SOLUTIONS

The People







An informal workplace to stimulate free expression and bring out personal skills.

Roma
Via Giovanni Squarcina, 3 - 00143 Roma
Tel: +39 06 5032715

Rende
Via Pedro Alvares Cabral, 6 - 87036 Rende (CS)
Tel: +39 0984 395257

Pisa
Via Forti Umberto, 1 - 56121 Pisa (PI)
Tel: +39 050 9655012

Milano
Corso Sempione, 62 - 20149 Milano (MI)
Tel: +39 02 316282

integris

ICT & COGNITIVE SOLUTIONS

Make your decisions simpler providing disruptive Data Analytics, Cognitive and Predictive Solutions, that allow you to explore "complex constellations of data" for getting out important information and gain competitive advantage.
This is our mission!

Our Solutions

Integris Explora ™ is the Data Analytics and Cognitive Computing solution designed and developed by Integris Spa. It enables the collection and analysis of structure, unstructured and multimedia data to discover their value, rules, trends and correlations and obtain important information to support decisions

explora TEXT

Explora Text: Solution for the automatic text analysis from online and/or offline sources based on Natural Language Processing

explora PROCESS

Explora Process: Process mining solution for the process analysis, discovery and enhancement. It reveals inefficiencies, bottlenecks and paths of executions, supporting process optimization.

explora SPEECH

Explora Speech: Solution for speech transcription and analysis of digital multimedia contents.

Core Business Areas

(Big) Data Analytics

- Descriptive / Predictive Analytics
- Customer Behaviour Analysis
- Cross Selling / Upselling Analysis
- Outlier Detection
- Predictive Maintenance
- Customer Experience Management
- Process Intelligence / Mining
- IoT Analytics

Cognitive Computing

- Sentiment Analysis
- Brand Reputation Analysis
- Information Extraction
- Semantic / Intelligent Search
- Media Monitoring
- Competitive Analysis
- Documental Classification (semantic / automatic)
- BOT / Service Automation
- Conversational AI / Digital Assistant

ICT Services

- IT Solutions (Design and Implementation)
- System Integration
- IT / Applications Modernization
- Support / Maintenance Services
- Microsoft, Hitachi, Pentaho, IBM, Tibco, Oracle, Splunk

Roma
Via Giovanni Squarcina, 3 - 00143 Roma
Tel: +39 06 5032715

Rende
Via Pedro Alvares Cabral, 6 - 87036 Rende (CS)
Tel: +39 0984 395257

Pisa
Via Forti Umberto, 1 - 56121 Pisa (PI)
Tel: +39 050 9655012

Milano
Corso Sempione, 62 - 20149 Milano (MI)
Tel: +39 02 316282

integris

ICT & COGNITIVE SOLUTIONS

Make your decisions simpler providing disruptive Data Analytics, Cognitive and Predictive Solutions, that allow you to explore "complex constellations of data" for getting out important information and gain competitive advantage.
This is our mission!

Key Technologies

Data Viz

Power BI · pentaho · tableau · Qlik · Superset

Data Analysis

R · python · MATLAB · Open for Innovation KNIME · sas THE POWER TO KNOW.

Data Eng

Microsoft Azure · HORTONWORKS hadoop · pentaho · talend · APACHE Spark · Greenplum · MySQL

Partnerships & Agreements

UNIVERSITÀ DI PISA · SCUOLA NORMALE SUPERIORE · UNIVERSITÀ DEGLI STUDI DI MILANO · UNIVERSITÀ DEGLI STUDI DI SALERNO CAMPUS VIVENDI · UNIVERSITÀ DELLA CALABRIA · UNIVERSITÀ DEGLI STUDI DI TRIESTE · UNIVERSITAS STUDIORUM CASSINAS · UNIVERSITÀ DI TRENTO · ALMA MATER STUDIORUM UNIVERSITÀ DI BOLOGNA · UNIVERSITÀ DI SIENA 1240 · POLITECNICO DI TORINO · UNIMORE UNIVERSITÀ DEGLI STUDI DI MODENA E REGGIO EMILIA · SAPIENZA UNIVERSITÀ DI ROMA · POLITECNICO MILANO 1863 · UNIVERSITÀ DEGLI STUDI FIRENZE · UNIVERSITÀ DEGLI STUDI DI NAPOLI FEDERICO II · ROMA TRE UNIVERSITÀ DEGLI STUDI · Università di Roma Tor Vergata · UNIVERSITÀ DI PAVIA · Consiglio Nazionale delle Ricerche

Roma
Via Giovanni Squarcina, 3 - 00143 Roma
Tel: +39 06 5032715

Rende
Via Pedro Alvares Cabral, 6 - 87036 Rende (CS)
Tel: +39 0984 395257

Pisa
Via Forti Umberto, 1 - 56121 Pisa (PI)
Tel: +39 050 9655012

Milano
Corso Sempione, 62 - 20149 Milano (MI)
Tel: +39 02 316282

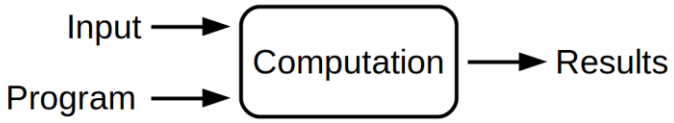
integris

ICT & COGNITIVE SOLUTIONS

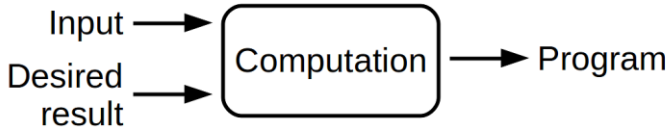
Machine Learning

The Vision

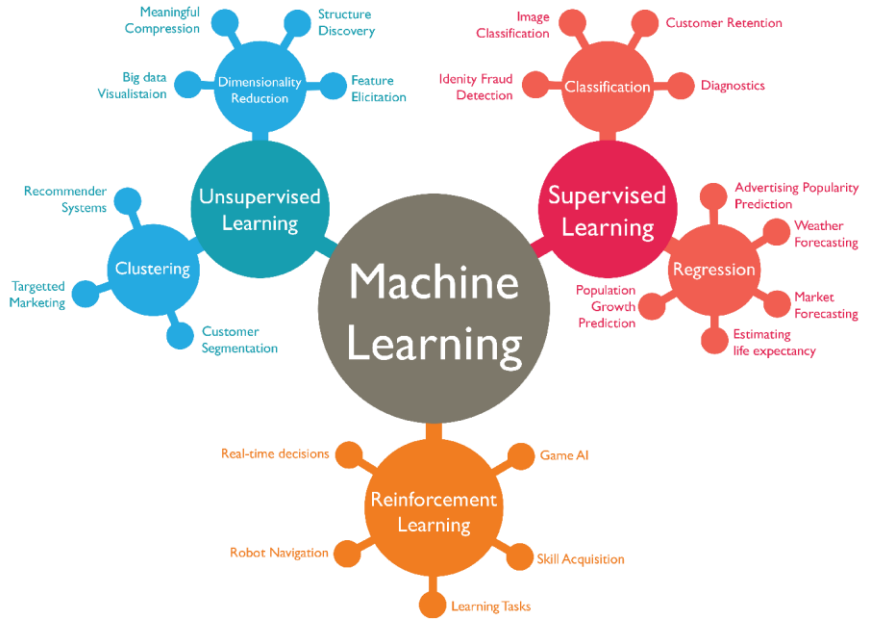
Traditional programming



Machine learning



The Techniques



Main Languages

python™

julia

Roma
Via Giovanni Squarcina, 3 - 00143 Roma
Tel: +39 06 5032715

Rende
Via Pedro Alvares Cabral, 6 - 87036 Rende (CS)
Tel: +39 0984 395257

Pisa
Via Forti Umberto, 1 - 56121 Pisa (PI)
Tel: +39 050 9655012

Milano
Corso Sempione, 62 - 20149 Milano (MI)
Tel: +39 02 316282

integris

ICT & COGNITIVE SOLUTIONS

Rstudio::sparklyr

- An R interface for Apache Spark
- Connect to Spark from R.
- The sparklyr package provides a complete dplyr backend.
- Using Spark SQL.
- Filter and aggregate Spark datasets then bring them into R for analysis and visualization.
- Use Spark's distributed machine learning library from R.
- Create extensions that call the full Spark API and provide interfaces to Spark packages.
- Pipelines to production with mleap

Roma
Via Giovanni Squarcina, 3 - 00143 Roma
Tel: +39 06 5032715

Rende
Via Pedro Alvares Cabral, 6 - 87036 Rende (CS)
Tel: +39 0984 395257

Pisa
Via Forti Umberto, 1 - 56121 Pisa (PI)
Tel: +39 050 9655012

Milano
Corso Sempione, 62 - 20149 Milano (MI)
Tel: +39 02 316282

integris

ICT & COGNITIVE SOLUTIONS

sparklyr Vs. sparkR

Feature	SparkR	sparklyr
Data input & output	+ +	+ +
Data manipulation	-	+ + +
Documentation	+ +	+ +
Ease of setup	+ +	+ +
Function naming	- -	+ + +
Installation	+	+ +
Machine learning	+	+ +
Range of functions	+ + +	+ +
Running arbitrary code	+	+ +
Tidyverse compatability	- - -	+ + +

https://eddjberry.netlify.com/post/2017-12-05-sparkr-vs-sparklyr/

Roma
Via Giovanni Squarcina, 3 - 00143 Roma
Tel: +39 06 5032715

Rende
Via Pedro Alvares Cabral, 6 - 87036 Rende (CS)
Tel: +39 0984 395257

Pisa
Via Forti Umberto, 1 - 56121 Pisa (PI)
Tel: +39 050 9655012

Milano
Corso Sempione, 62 - 20149 Milano (MI)
Tel: +39 02 316282

integris

ICT & COGNITIVE SOLUTIONS

sparklyr::main functions

Three families of functions:

- Machine learning algorithms for analyzing data (ml_*)
- Feature transformers for manipulating individual features (ft_*)
- Functions for manipulating Spark DataFrames (sdf_*)

DEMO

install locally, no cluster, and open connection

spark_install('2.0.1')

sc <- spark_connect(master='local')

Copy data to Spark memory: *x <- copy_to(sc, iris, 'spark_iris', overwrite=TRUE)*

Partition data: *prt_x <- sdf_partition(x, training = 0.7, test = 0.3)*

Create table in R and use dplyr: *tbl(sc, 'prt_x') %>% select(Species, Petal_Length)*

https://www.rstudio.com/resources/cheatsheets/#sparklyr

Roma
Via Giovanni Squarcina, 3 - 00143 Roma
Tel: +39 06 5032715

Rende
Via Pedro Alvares Cabral, 6 - 87036 Rende (CS)
Tel: +39 0984 395257

Pisa
Via Forti Umberto, 1 - 56121 Pisa (PI)
Tel: +39 050 9655012

Milano
Corso Sempione, 62 - 20149 Milano (MI)
Tel: +39 02 316282

integris

ICT & COGNITIVE SOLUTIONS

sparklyr::MLlib

MLlib is Apache Spark's scalable machine learning library usable in Java, Scala, Python and R.
MLlib contains many algorithms about:

- Classification and Regression (decision tree, logistic regression, ...)
- Clustering (k – means, LDA, ...)
- Collaborative Filtering ()
- Pattern Mining

In our project we used two **Ensemble Algorithms**:

- Random Forest
- Gradient Boosting

Roma
Via Giovanni Squarcina, 3 - 00143 Roma
Tel: +39 06 5032715

Rende
Via Pedro Alvares Cabral, 6 - 87036 Rende (CS)
Tel: +39 0984 395257

Pisa
Via Forti Umberto, 1 - 56121 Pisa (PI)
Tel: +39 050 9655012

Milano
Corso Sempione, 62 - 20149 Milano (MI)
Tel: +39 02 316282

integris

ICT & COGNITIVE SOLUTIONS

sparklyr::deploy!

- Create a pipeline with Mleap
- Use RStudio server
- Deploy in another language (our case):
  - PoC in R with Sparklyr
  - Deploy in production using Java API for MLlib

Roma
Via Giovanni Squarcina, 3 - 00143 Roma
Tel: +39 06 5032715

Rende
Via Pedro Alvares Cabral, 6 - 87036 Rende (CS)
Tel: +39 0984 395257

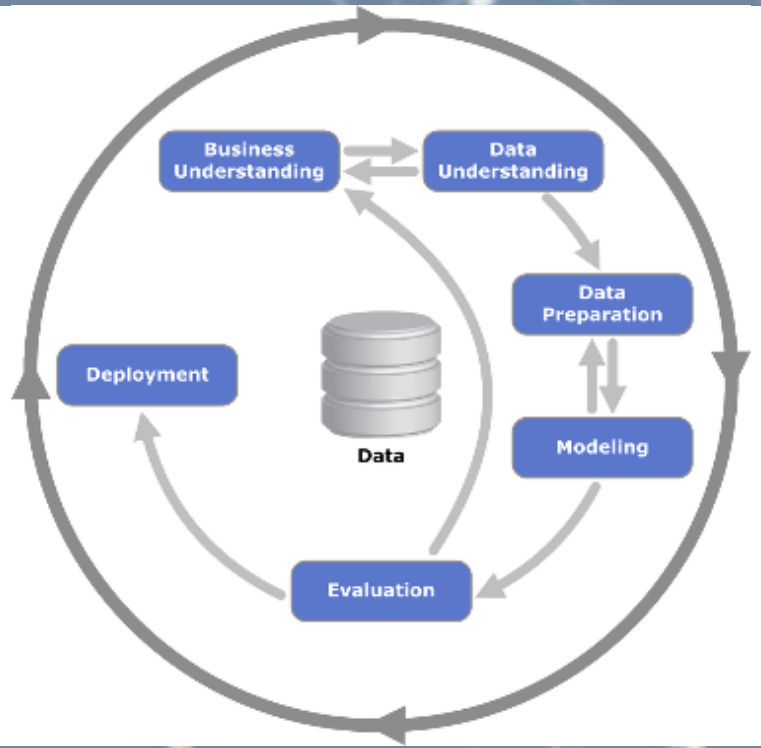
Pisa
Via Forti Umberto, 1 - 56121 Pisa (PI)
Tel: +39 050 9655012

Milano
Corso Sempione, 62 - 20149 Milano (MI)
Tel: +39 02 316282

integris

ICT & COGNITIVE SOLUTIONS

CRISP - DM



Cross-industry standard process for data mining is an open standard process model that describes common approaches used by data mining experts.

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Modeling
5. Evaluation
6. Deployment

Roma
Via Giovanni Squarcina, 3 - 00143 Roma
Tel: +39 06 5032715

Rende
Via Pedro Alvares Cabral, 6 - 87036 Rende (CS)
Tel: +39 0984 395257

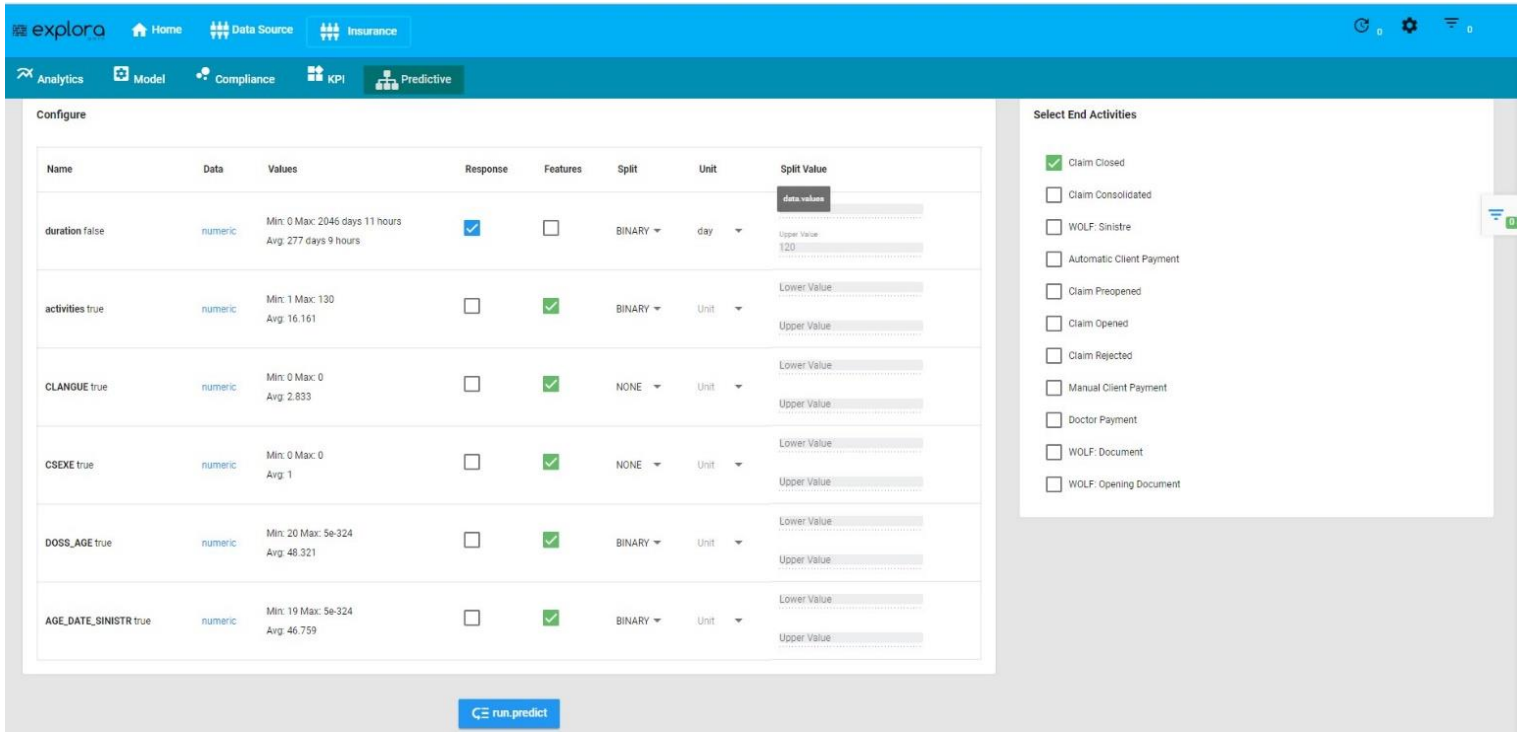
Pisa
Via Forti Umberto, 1 - 56121 Pisa (PI)
Tel: +39 050 9655012

Milano
Corso Sempione, 62 - 20149 Milano (MI)
Tel: +39 02 316282

integris

ICT & COGNITIVE SOLUTIONS

Explora Process



Select:

- Features
- Response

Roma
Via Giovanni Squarcina, 3 - 00143 Roma
Tel: +39 06 5032715

Rende
Via Pedro Alvares Cabral, 6 - 87036 Rende (CS)
Tel: +39 0984 395257

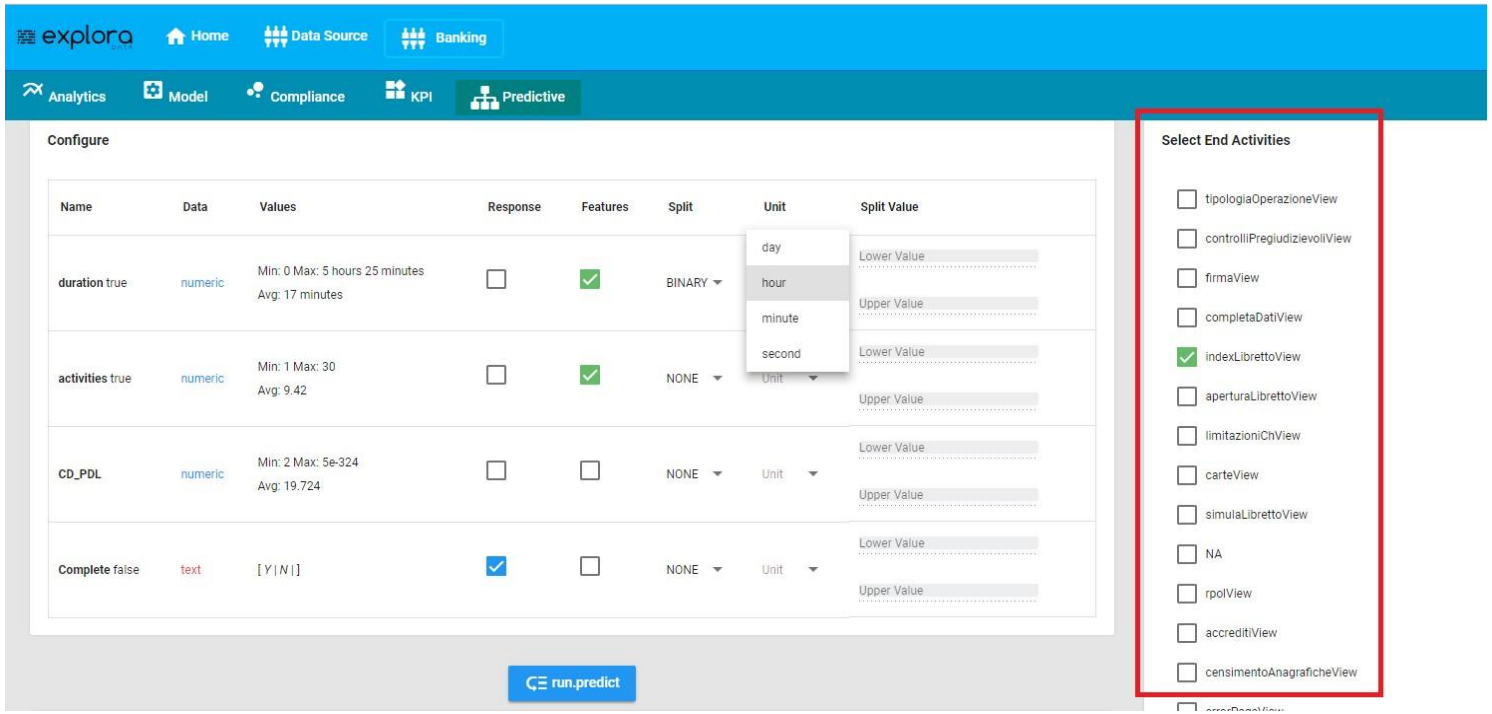
Pisa
Via Forti Umberto, 1 - 56121 Pisa (PI)
Tel: +39 050 9655012

Milano
Corso Sempione, 62 - 20149 Milano (MI)
Tel: +39 02 316282

integris

ICT & COGNITIVE SOLUTIONS

Explora Process



Select:

- End Activity: split the dataset in train (closed cases) and test set (open cases).
- Time granularity

Roma
Via Giovanni Squarcina, 3 - 00143 Roma
Tel: +39 06 5032715

Rende
Via Pedro Alvares Cabral, 6 - 87036 Rende (CS)
Tel: +39 0984 395257

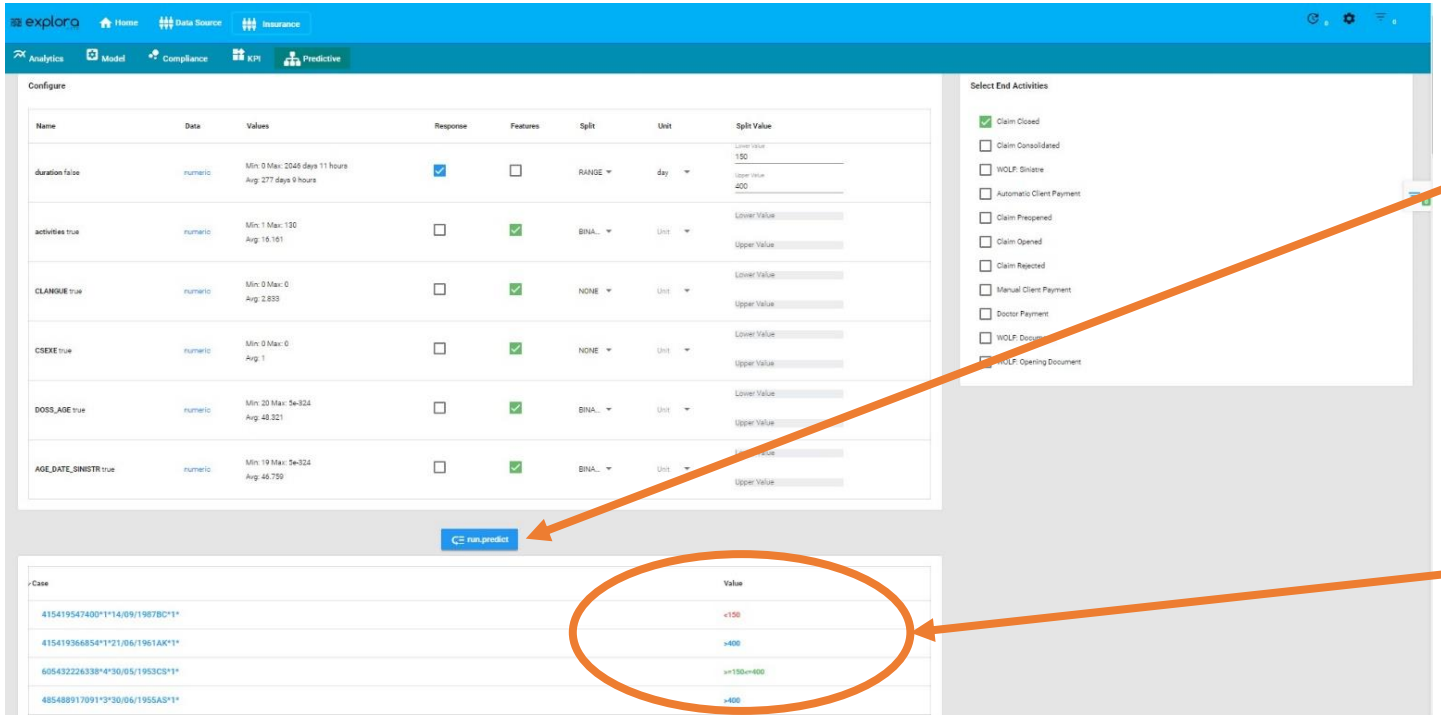
Pisa
Via Forti Umberto, 1 - 56121 Pisa (PI)
Tel: +39 050 9655012

Milano
Corso Sempione, 62 - 20149 Milano (MI)
Tel: +39 02 316282

integris

ICT & COGNITIVE SOLUTIONS

Explora Process



RUN

PREDICT

Roma
Via Giovanni Squarcina, 3 - 00143 Roma
Tel: +39 06 5032715

Rende
Via Pedro Alvares Cabral, 6 - 87036 Rende (CS)
Tel: +39 0984 395257

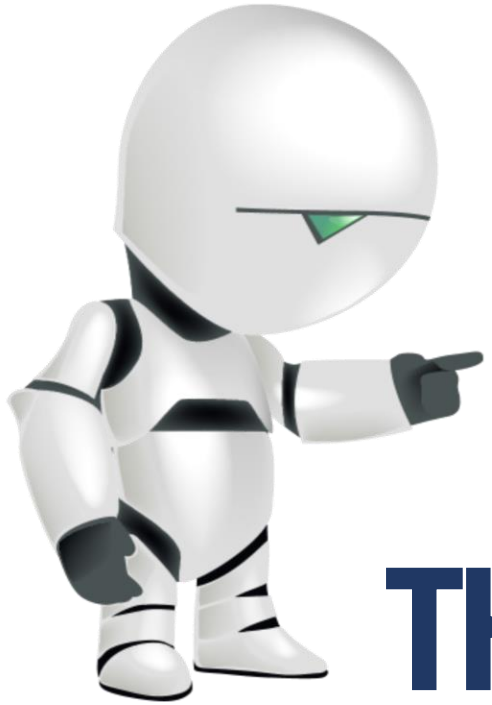
Pisa
Via Forti Umberto, 1 - 56121 Pisa (PI)
Tel: +39 050 9655012

Milano
Corso Sempione, 62 - 20149 Milano (MI)
Tel: +39 02 316282

integris

ICT & COGNITIVE SOLUTIONS

Q&A



nicola.procopio@integris.it

Nicola Procopio

nickprock

THANKS AND KEEP IN TOUCH!

Roma
Via Giovanni Squarcina, 3 - 00143 Roma
Tel: +39 06 5032715

Rende
Via Pedro Alvares Cabral, 6 - 87036 Rende (CS)
Tel: +39 0984 395257

Pisa
Via Forti Umberto, 1 - 56121 Pisa (PI)
Tel: +39 050 9655012

Milano
Corso Sempione, 62 - 20149 Milano (MI)
Tel: +39 02 316282