








Data & Analytics applications in Greek insurance market  
**Interamerican** case study

# Introductions

## Who I am



## Current info's

-  **Babis Babilis**
-  **Data & Analytics Leader**
-  **Interamerican, part of Achmea**
-  /babisbabilis
-  mpampilisch@interamerican.gr

## Studies

-  **MSc Actuarial Science & Risk Management**  
University of Piraeus, dep: Statistics, (2013-2015)
-  **BSc Applied Economics**  
HUA University, dep: Economics, (2007-2012)









## Professional Experience

-  **Math's Tutor, (2013-2014)**
-  **Non-Life Pricing Actuary, (2014-2015), IAG**
-  **P&C Data Analyst, (2015-2017), IAG**
-  **Data Scientist, (2017-2019), IAG**
-  **Data & Analytics Leader, (2019- ), IAG**

## Certifications

-  **Executive Data Science**  
John Hopkins University, (2019)
-  **Analytical Accelerator Program**  
MIcompany, Netherlands, (2016)

## Skills / Expertise

-  **Leadership** 80%  

-  **Predictive modeling / ML** 90%  

-  **Dynamic Pricing** 90%  

-  **Data Engineering / DW** 70%  


**Interamerican** use the power and the insights of the data for supporting the customer needs and expectations by creating better services and products

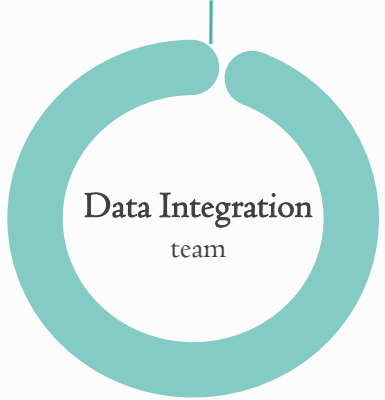
### Data & Analytics Team

Create **architectural databases** that can provide **storage, use, normalization of data** based on **business & analytics needs** with ETL or ELT processes through different source systems

Using **data** in order to create **predictive modeling algorithms, machine learning technics & artificial intelligence** solutions helping the company shifting in more advanced analytics solutions

Manage the **accuracy, validity, timeliness, completeness, uniqueness and consistency** of the data in systems and data flows using **methodical approaches, policies and processes**

Using **data** in order to **analyze, monitor, understand and serve** business needs through data analysis solutions



Data Integration team



Data Science team



Data Governance team



Data Analysis team



Data Engineer



Data Scientists



Data Analyst



Data Analyst

# Data Science implementations affects all business processes providing data driven decisions

Using **data** in order to create **predictive modeling algorithms, machine learning technics & artificial intelligence** solutions helping the company shifting in more advanced analytics solutions



**Customer Satisfaction**

- Customer acquisition (Machine learning, Predictive analytics)
- Customer retention (Machine learning, Predictive analytics)
- Customer segmentation (Predictive analytics)
- Customer personalization (Machine learning, Predictive analytics)

**Operational Efficiency**

- Cost reductions (Predictive analytics, Data engineering)
- Processes automation (Data engineering)
- Risk assessment (Predictive analytics)

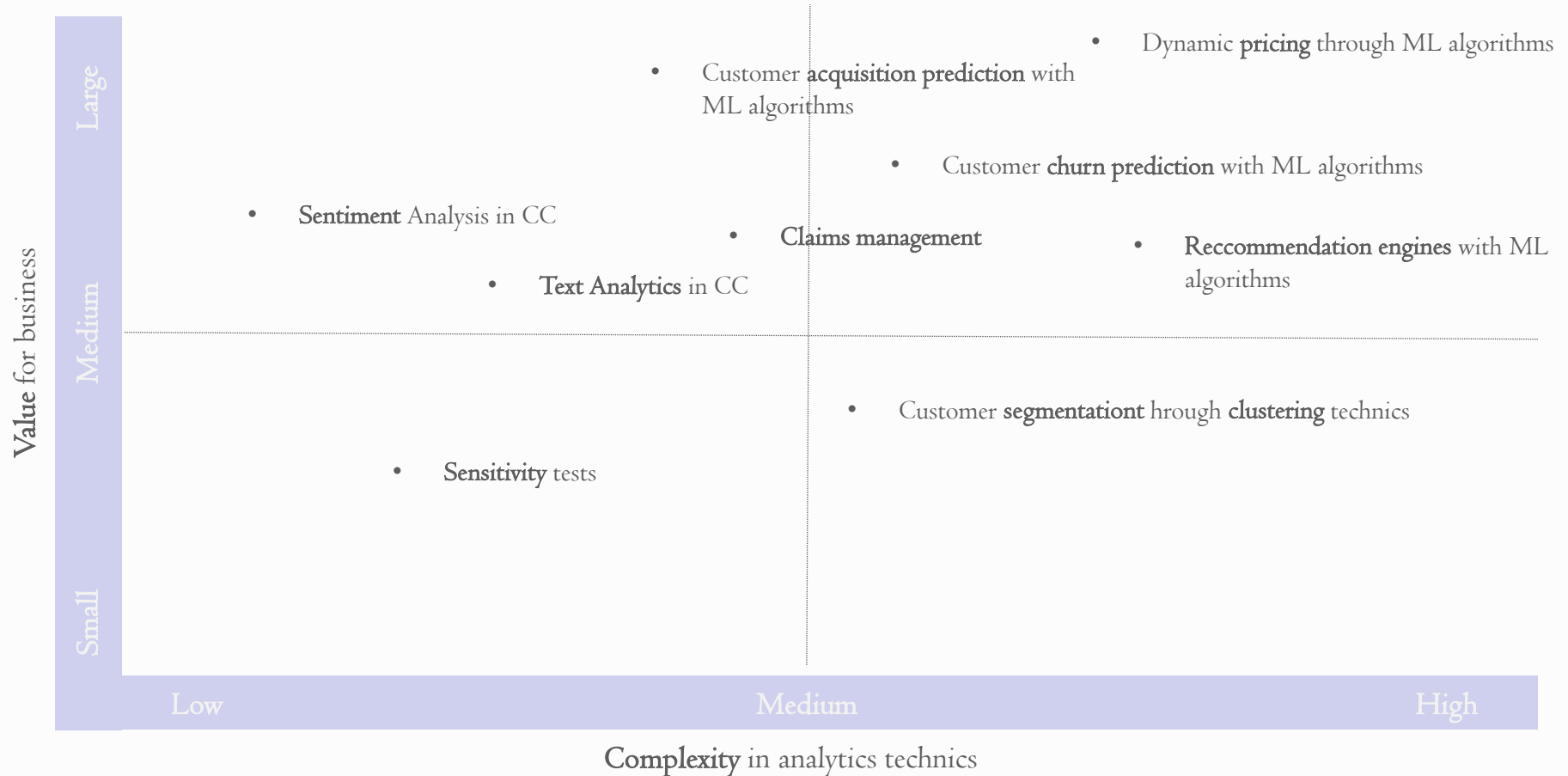
**Claims Management**

- Fraud detection (Machine learning)
- Fraud prevention (Machine learning)
- Processes automation (Data engineering, Predictive analytics)

**Profitable Growth**

- Pricing optimization (Machine learning, Predictive analytics)
- Profitability optimization (Data engineering, Predictive analytics)
- Portfolio management (Predictive analytics)
- Lifetime value prediction (Machine learning, Predictive analytics)

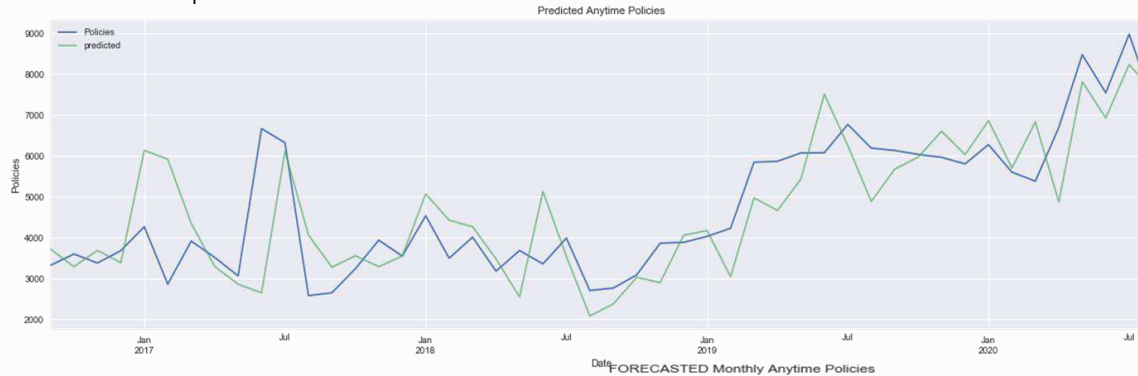
- Machine learning
- Predictive analytics
- Data engineering



# Case Study I - Forecasting

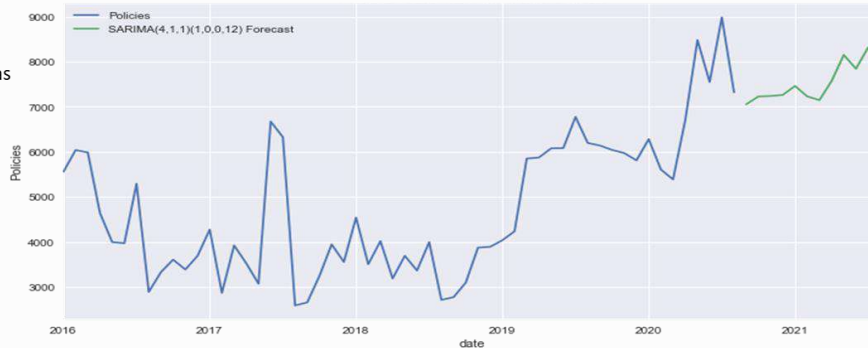
Forecasting of NB policies & demand provides with better understanding on one-step-ahead conversion and identification of areas of improvement

A **SARIMA(4, 1, 1) (1, 0, 0, 12)** model was fit on **Monthly Policies** data. The following graph compares the actual Timeseries with the predicted one.



- Every quarter re-run of the models should be done for better accuracy

- ✓ The forecasted Policies timeseries follows the increasing trend that was visible during the last months, along with the monthly Seasonality.
- ✓ Data was stationarized with First order Differentiation (d=1) in order to proceed with optimal model fit.

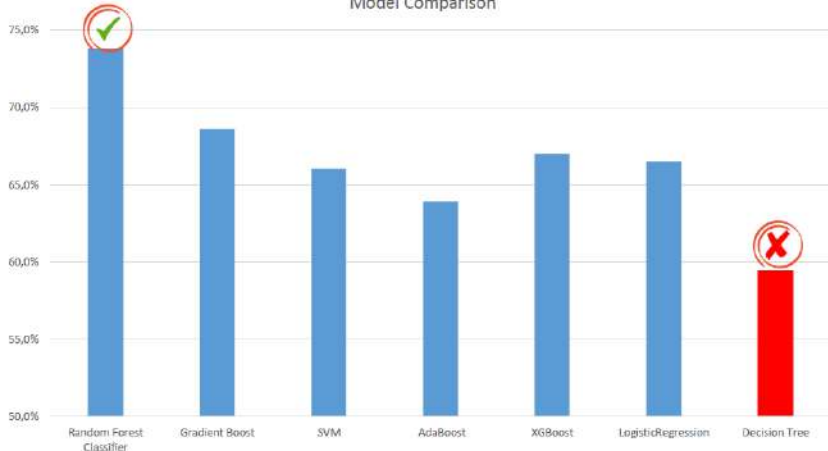


# Case Study 2 – Recommendation Model

Recommendation engine ML algorithm provides sophisticated customer personalization on Anytime website



Model Comparison



Random Forest Model		Total Average Accuracy		
Suggested Plan (down)	Plan that the customer bought (right)	Basic	Premium	Value
BASIC		67.4%	0.0%	32.6%
PREMIUM		5.4%	4.2%	90.5%
VALUE		20.1%	0.0%	79.9%
<b>Grand Total</b>		<b>41.9%</b>	<b>0.1%</b>	<b>58.0%</b>

Decision Tree		Total Average Accuracy		
Suggested Plan (down)	Plan that the customer bought (right)	Basic	Premium	Value
BASIC		59.62%	0.12%	40.26%
PREMIUM		22.91%	10.39%	66.70%
VALUE		32.73%	2.79%	64.48%
<b>Grand Total</b>		<b>48.72%</b>	<b>1.46%</b>	<b>49.82%</b>

# Case Study 3 – Churn predictive model

The churn predictive model expected probabilities per customer combined with Loss Ratio results is used for better retention management actions from Interamerican

## Models

**Logistic Regression**



**Boosting**



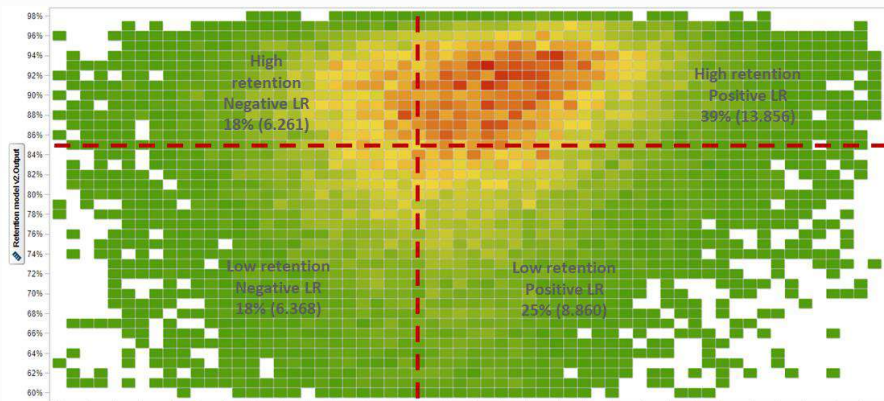
**Random Forest**





Ada Boost	XGBoost	Cat Boost
<ul style="list-style-type: none"> <li>Each predictor pays more attention to the instances wrongly predicted by its predecessor.</li> <li>Achieved by changing the weights of training instances</li> </ul>	<ul style="list-style-type: none"> <li>Sequential correction of predecessor's errors.</li> <li>Does not tweak the weights of training instances.</li> <li>Fit each predictor is trained using its predecessor's residual errors as labels</li> <li>Missing Values handling</li> <li>Prevent Overfitting</li> <li>Fast Performance</li> </ul>	<ul style="list-style-type: none"> <li>Similar to XGBoost</li> <li>Categorical Feature handling</li> <li>Missing values Handling</li> <li>Avoid overfitting</li> <li>Excellent default hyperparameter results (less tuning)</li> <li>Create symmetric trees</li> <li>Even fast performance than XGBoost</li> </ul>

## Results

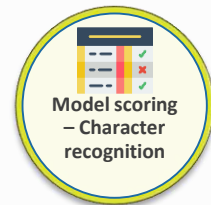
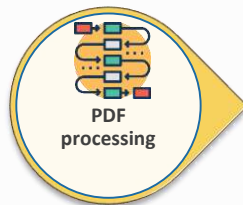
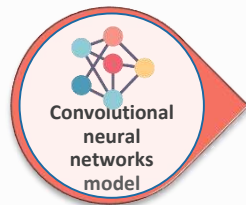


- The above chart represents the overall portfolio scored with ML algorithms for churn probabilities and combined with Loss Ratio results per customer creates four potential pool of actions based on the budgeted retention rates & LR numbers.
- Red dots are the highly concentration of customers
- Higher retention (lower churn) together with lower LR is the most strategic choice for the portfolio optimization



# Case Study 4 – Artificial Intelligence in handwritten pdfs

This implementation support the broader digitalization process of the company



### 3 datasets of 485k images of:

- Greek characters \*
- English characters
- Digits

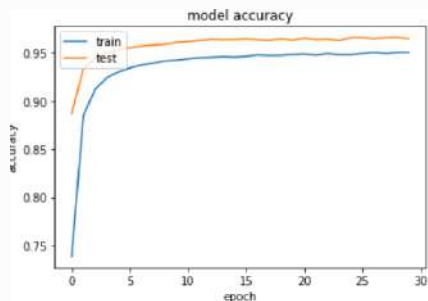
### transformed to:

- Images of size 28X28
- array of pixels (1 = black pixel, 0 = white)
- labeled\classes

The convolutional neural network (CNN) is a deep learning neural network.

picture (28X28) --- CNN ---- class

### Accuracy tests



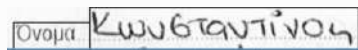
### Page skewness

- Correct image skew (if any)

### Page noise removal

- Remove colors
- Remove lines

### Text line & Word detection



### Character detection

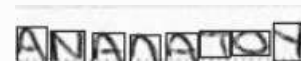
- Split each word to characters

### Character binarization

- Convert pixels to black and white

### Character resizing

- Resize to 28 X 28 pixels



### Model scoring

- Score each character

### Result manipulation

- Connect characters to form words
- Make it iterative

Thank You / Questions