Customs & Documents Compliance Innovation Initiatives
Workshop program

15.30 - 16.30  Customs & Documents Innovation Initiatives
16.30 - 16.45  Coffee Break
16.45 - 17.30  Interactive workshop
17.30 - 18.00  Results & Recap
“You can have data without information, but you cannot have information without data.”

- Daniel Keys Moran

“Data is not information, information is not knowledge, knowledge is not understanding, understanding is not wisdom.”

- Clifford Stoll
Industry 1 to 4.0 SMART - Digitalization
Customs Initiatives (NL)

Pushing Boundaries
- Controls out of the border
- System based approach
- Coordinated border management

Keyword
- Auto detection (data/goods)
<table>
<thead>
<tr>
<th>Data</th>
<th>Goods</th>
</tr>
</thead>
<tbody>
<tr>
<td>Automated creation of signals and insights utilizing Big Data in order to enforce and ensure customs compliance.</td>
<td>Automated interpretation of signals derived from scan and/or detection technologies.</td>
</tr>
</tbody>
</table>
Apply algorithms
Improve clearance data by enrichment (free/public data sources)
Improve clearance data by enrichment (protected data sources)
Signals and insights from scans/detection goods
Use data pipelines
Apply filter methodology
Apply data analytics
Create data visualization

Trusted Trader
Trusted Trade Lanes
Other
DATA

*a set of values of subjects with respect to qualitative or quantitative variables*
From data to knowledge

DATA
Interpretation
Analytics

INFORMATION

Practice
Education

KNOWLEDGE

Change
Correct

Abstract
Data becomes information

- Collected
- Measured
- Analyzed
- Reported
- Enriched

Visualization
Data correction
Outliers
Data entry errors
Top 5 existing HS codes ship to

- China: 5,5
- Japan: 4,3
- USA: 4,1
- Egypt: 2,3
- UAE: 0,8
Auto detection of data
Relevant signals and alerts

- HS codes
- Origin
- Valuation
- Dual Use codes
- Incoterms
- Weight and dimensions
- AEO number

- EORI identification
- Description of merchandise
- Document codes
- Certificates and licenses
- Duty rates
- And so on...
Scan innovation

TCM app Customs

- Alert: contains illegal ingredients!
- No hits found
Documents innovations

PINCVISION COLLABORATES WITH THE AWTCC
22 Jul 2019 at 12:00 am

In order to increase Pincvision’s independence and to strengthen our global reach, Pincvision has signed a partnership with the American World Trade Chamber of Commerce (AWTCC).

Thanks to this innovative collaboration, Pincvision can now also create legalized export documents for customers who:

- Ship from the United States
- Ship goods with US origin
- Are part of a US entity
Documents innovations

THE PAPER ATA CARNET IS GOING DIGITAL

20 Jun 2019 at 12:00 am

Did you already know that the ATA Carnet will enter the digital world? The system is ready and the International Chamber of Commerce (ICC) will start with the first test phase.

A nice step forward in the ATA Digitization project. Earlier this year, we announced that this digitization project was going to start this summer.

We are very enthusiastic that the International Chamber of Commerce (ICC) now stated that the technical system is ready and the pilot running phase is about to start. They also posted a short movie on LinkedIn about this pilot to inform you about the steps that need to be taken in this digitization project.
Pincvision’s Customs Innovations

1. Master data validation and enrichment
2. Data assessment and visualization
3. Automated HS classification
Pincvision’s Customs Innovations

Example Master Data Validation and Enrichment

Input: HS codes + Country of Origin

Enrich
Pincvision’s Customs Innovations

- Data assessment and visualization
Pincvision’s Customs Innovations

- Automated HS classification
  
  Basis of auto HS = register of already classified articles
  
  Threshold: 90% similarity

- 90%-100% based on attributes/features
- <90% based on attributes/features
- Outliers
The examples of the authorities discussed before have something in common. They are using Artificial Intelligence (AI) algorithms.

But what are those AI algorithms? Let’s take a closer look at those algorithms.

What is it?
Where are they used from a business perspective?
Which groups of algorithms are available?
How can we use them?

But let’s start with some definitions to have some context.
AI is about developing computer systems to perform tasks requiring human intelligence. Definition on quite a high level. But AI can be divided in multiple subgroups like machine learning and neural networks.

When we look at machine learning it’s already more specific. But we start with the basic definition of a algorithm.

Machine Learning is made up of a series of algorithms. Basically, AI (Machine Learning is a subset of AI) is designed to learn in the same way as a child. Thanks to a dataset, an AI can find patterns and builds assumptions based on those findings. This is called model-based learning, and it allows AI to make better decisions than humans because it can take many more factors into account and analyze them in milliseconds.

An algorithm is like following a recipe. You follow a set of instructions: prepare the ingredients, heat the oven to 200C, and bake for 10 minutes. The output/result will be to have a great cake.

Now, let’s imagine your oven is too hot. Through Machine Learning, the system learns from the past that the oven gets too hot and so turns it down.
**Algorithm**

a process or set of rules to be followed in calculations or other problem-solving operations, especially by a computer. Basically, the goal of an algorithm is to solve a specific problem, usually defined by someone as a sequence of steps.

**Machine learning**

set of algorithms that enable the software to update and “learn” from previous outcomes without the need for programmer intervention. It is fed with structured data in order to complete a task without being programmed how to do so.

Machine Learning can be compared with an experienced cook (if you have a good dataset). It knows the recipe and has learned a lot from previous experiences. For instance, the system has found out that this ingredient worked well with this cake (based on data) and it would make recommendations/predictions. By using ML, you get something that goes beyond the sum of its parts.
In the beginning of AI/machine learning it was more the science world that was using it and also made big steps in the development and usage. When this “new world” was coming to the business it was e.g. called data mining and used most in marketing. Questions in this area were often sales related.
Sales forecasting
Segmentation/Clustering of customers
Find relationships between products

Nowadays a big industry is gaming and call systems (website)

When we look at trade compliance you can think about Detecting/preventing fraud, autofill missing data elements and anomaly detection. The last two will be discussed in more detail later on

Can be used anywhere...
Big changers
- Maturity of technology & amounts of data

Why is it now more usable for end-users?
- Lot of “auto” tuning inside the services
- Auto morphing of algorithm variants

Scientist had a problem. They have knowledge of all those algorithms but they had a lot of data
They where responsible for a big game changer. Data-drive approach

As technology is further improving it’s nowadays a sort of service
Morphing of usage (variant)
“Autotune” based of feedback user is giving to the system

Why is it now more usable for end-users?
Lot of “auto” tuning inside the services itself
Auto morphing of algorithm variants (use the best variant for your question)
Almost like buying a car.. (don’t understand everything under the hood but you can drive it)
1985 — Terry Sejnowski invents NetTalk, which learns to pronounce words the same way a baby does.
1990s — Work on machine learning shifts from a knowledge-driven approach to a data-driven approach. Scientists begin creating programs for computers to analyze large amounts of data and draw conclusions or “learn” from the results.
1997 — IBM’s Deep Blue beats the world champion at chess.
2006 — Geoffrey Hinton coins the term “deep learning” to explain new algorithms that let computers “see” and distinguish objects and text in images and videos.
2010 — The Microsoft MSFT +0% Kinect can track 20 human features at a rate of 30 times per second, allowing people to interact with the computer via movements and gestures.
2011 — IBM’s Watson beats its human competitors at Jeopardy.
2011 — Google GOOGL +0% Brain is developed, and its deep neural network can learn to discover and categorize objects much the way a cat does.
2012 – Google’s X Lab develops a machine learning algorithm that is able to autonomously browse YouTube videos to identify the videos that contain cats.
2014 – Facebook FB +0% develops DeepFace, a software algorithm that is able to recognize or verify individuals on photos to the same level as humans can.
2015 – Amazon launches its own machine learning platform.
2015 – Microsoft creates the Distributed Machine Learning Toolkit, which enables
the efficient distribution of machine learning problems across multiple computers. 2015 – Over 3,000 AI and Robotics researchers, endorsed by Stephen Hawking, Elon Musk and Steve Wozniak (among many others), sign an open letter warning of the danger of autonomous weapons which select and engage targets without human intervention.

2016 – Google’s artificial intelligence algorithm beats a professional player at the Chinese board game Go, which is considered the world’s most complex board game and is many times harder than chess. The AlphaGo algorithm developed by Google DeepMind managed to win five games out of five in the Go competition.

Ethics and the physiologic side is more and more important. What is allowed? Like the DNA question Weapons, crumpy “agents” and systems that developed their own language to communicate
Start with a project but then from that moment it’s a business process
1. **Business understanding**
   Knowledge of your processes, what is the goal
2. **Data understanding**
   Definitions of data
   Data quality
3. **Data collection and preparation**: everything from choosing where to get the data, up to the point it is clean and ready for feature selection/engineering. First, thee take about 80% of the time this includes all changes to the data from once it has been cleaned up to when it is ingested into the machine learning model split in training set and test set
4. **Choosing the machine learning algorithm and training our first model**: getting a "better than baseline" result upon which we can (hopefully) improve

**Supervised learning**
Data including answer (label)

**Unsupervised**
Only data, algorithm should detect the value itself

**Semi supervised**
Small data supervised (with label), rest unlabeled

**Re-enforced learning**
Used can give feedback (punish/ treat)

**Evaluating our model:** this includes the selection of the measure as well as the actual evaluation; seemingly a smaller step than others, but important to our end result

**Model tweaking, regularization, and hyperparameter tuning:** this is where we iteratively go from a "good enough" model to our best effort
Forecasting
Analyzes and forecasts time-based data, combining the power of short-term predictions and long-term accuracy algorithms.

Regression (Supervised Linear and logistic)
Determines the relationship between columns in order to predict/evaluate an outcome (probability that a column will contain a specific state)
More usable with *numerical* values like e.g. amounts and kilograms

Clustering (unsupervised)
Classifies cases into distinctive groups based on any attribute sets

Association (unsupervised)
Identifies relationships between cases (e.g. webshops)

Classification (supervised)
Classifies cases into distinctive groups based on any attribute sets (predefined classes). Different learning technic compared to clustering

**Supervised learning**
Data including answer (label)

**Unsupervised**
Only data, algorithm should detect the value itself

**Semi supervised**
Small data supervised (with label), rest unlabeled

**Re-enforced learning**

Used can give feedback (punish/ treat)

Trail and Error, search for the biggest reward
Almost everybody is using this algorithm every day..

**Spam**
It’s your spam filter in the mail!!

60..75% of all mail is spam. Advanced algorithm compared to the beginning

**HS Classification**
Exception GIR = General Interpretation Rules

Case law jurisprudence

One step further, assume you have hundreds of new products a week that should be classified. How can we check them? Is there a way to select the suspicious ones?
We love to think that data is well behaved and sensible, unfortunately this is often not the case.
- Erroneous data points due to malfunctions or errors
- Missing data elements

Detecting anomalies provides a measure of quality control.

Sometimes anomalies are indicative of a real problem and are not easily explained.

Example: Toy boat (for kids) is classified as a yacht.
- Product description contain a “boat”
- HS Code is totally different compared to the other toys (product type)
- Product dimension, weight and values are totally different
A few slides ago you saw a high-level grouping of popular algorithms. But if you look closed there are many more variant. This is just a small list of variants. It will be worse because each implementation of a supplier is using one or more variants and giving them fancy names 😊 Does it matter? No just be aware of it when you take a closer look.

Most important is that you know what your product is using:
Algorithm categories
Learning method (supervised, unsupervised, semi-supervised..)
Backflipping Robot: https://www.youtube.com/watch?v=fRj34o4hN4I
Bike riding Robot: https://www.youtube.com/watch?v=mT3vfSQePcs
With all those algorithms we’re still generating new data or checking existing data but we want to change this to information to show share this with other people. Yuri used this graph a few slides back. He already mentioned that we could improve this visualization..

How to show the data in a few that its clear, efficient and easy to read.

Can we show a excel sheet to other people? Yeas we can, it this efficiently? No because how does the other person know where to look at?

We can use graphs to show specific data points ..
Some types can be used for multiple purposes
Choose is dependent on your goal (focus)

Colorblindness 8% of men and 0,5% of female, also when printing in Black/White

**Comparison**
Use these visuals when you want to display measures compared by its magnitude.

**Change over time**
Use these visuals when you want to display the changing trend of measures

**Ranking**
Use the visuals when you want to display measures by its rank order

**Spatial (Geo)**
Use these visuals when you want to display measures over spatial maps (Geo)

**Flow**
Use these visuals when you want to display flow or dynamic relations between measures

**Part to whole**
Use these visuals when you want to display parts that compose measures

**Distribution**
Use these visuals when you want to display the distribution of a measure

Correlation
Use these visuals when you want to display relations between measures

Single
Use these visuals when you want to display a single value

Tips
Does anybody know how many people are colorblind?
A Design Problem
This graph was found in Business Objects' user documentation. It served as an example of 3-D bar graphs. Business Objects is a leading business intelligence software vendor. Take a moment and try to interpret the data.

Analysis
Notice the following problems in the example above:
The 3-D bars are impossible to read.
The heavy grid lines offer nothing but distraction.
The vertically-oriented labels (i.e., the resort names and years) are difficult to read.
The years run from back to front, which is counter-intuitive.

Solution
This graph eliminates the use of 3-D altogether, which almost never works in business graphs. Also notice the other following improvements:
The bright bar colors have been replaced with variations in gray-scale, thereby guaranteeing that this graph will still work if it is photocopied or printed on a black-
and-white printer.
The three resorts have been arranged in order of rank, based on revenue, to highlight their comparative performance.
The years have been arranged from left to right, which is intuitive.
The legend has been placed above the bars, in the same sequence as the bars, to ease the process of matching them up.
The bar graph solution is ideal if people primarily want to focus on the magnitudes of a particular bar or if they want to compare the revenue of all three resorts in a given year. However, if we're more interested in seeing how revenue has changed through time, then the following line graph solution, suggested by Daniel Ferry, would work better.

This design makes it easy to see how revenue has changed from year to year at each of these resorts. The gray-scale color scheme was preserved that was already used in the bar graph, so this graph will also work well if printed in black-and-white. Additionally, because the labels were included of the lines directly the legend is removed.
In both graphs the exact same data is shown. Just see how much easier you can read the right graph.
Almost everybody is using dashboard nowadays. But what is a dashboard?
A dashboard is a tool used for information management. Much like the dashboard of a car, data dashboards organize, store, and display important information from (multiple) data sources into one, easy-to-access place to a specific audience. Using data visualization, dashboards uniquely communicate metrics visually to help users understand complex relationships in their data. In a data dashboard, it’s easier to draw parallels between different but related metrics, identify trends, and head off potential challenges hidden in an organization’s data.

‘A dashboard is a data visualization tool that displays on a single screen the status of business metrics, key performance indicators (KPIs) and important data points for an organization, department, team or process.’

Specific audience
(multiple) data sources
Often one page with max 7 till 9 items
data visualization (graphs, grids, KPI, ..)
head off potential challenges hidden in an organization’s data.
Let’s look at a brief introduction to these rules:

**Design for a target**
Simple, but underestimated: don’t lose sight of your goals and target users, otherwise you will fill the dashboard with too much useless information.

**Keep everything at a glance**
What if your car dashboard contained scrollbars? Ouch. A good dashboard must show all relevant information without the need for touch, scrolling, or clicks.
Guideline: 6 till 9 items per page

**Keep it simple**
Complexity could come across as sophistication, but in reality it is just an obstacle between your dashboard and the end users.

**Align elements**
By aligning the elements your page will be much cleaner. Not aligned page looks a bit messy..

**Be consistent**
Is repetition boring? Not really.

**Highlight the most relevant information**
Your dashboard is like a page of a magazine: each location has its meaning and a different level of importance. Don’t place charts at random.
We read from the left to the right and from up to down!. The place of the graph is
also important!

**Be clear**
Acronyms are bad. Legends are good.

**Start from zero**
Chart axes must be used consciously. Sometimes we convey the wrong message just by forgetting to pay attention to details.

**Shorten the numbers**
Dashboards users want to see the overall picture.

**Show the context**
Numbers only carry meaning within their context. Also supported by adding a good titles to the graph and axis, usage of legends

**Choose the right colors**
Do you know that about 10% of the population suffers from color blindness?
Printing black-White and colorblindness
Color associating, Red and Green..

**Design dashboards, not reports**
Remember, not all the details from a table are suitable for a dashboard.

**Show variations**
Don’t let users do the math.

**Leave the noise off**
Don’t suggest relations that don’t exist.

**Pick the right chart**
Each piece of information must be displayed using the right chart in your dashboard, so choose it wisely.

Specific audience
(multiple) data sources
Often one page with max 7 till 9 items
data visualization (graphs, grids, KPI, ..)
head off potential challenges hidden in an organization’s data.
At the back of the room there are numbers on the table. Please pick a number this will represent your group.
1. **Read case 1**
   What would you like to visualize based on the data (see datasheet)?
   - graphics/charts/tables/summaries?
     Please visualize

2. When we look at trade compliance based on Data and Data visualization, it will most likely include e.g. threats and weaknesses.
   - Please create a generic simplified SWOT with your group.