Machine Learning and Product Classification

Background

Artificial Intelligence (AI) is a term that's been used for decades. As far back as the 1950's science fiction shows frequently referred to it, even though it was little more than a concept at the time. Years later, IBM challenged the world's greatest chess player to a match against it's computer's artificial intelligence. An impressive leap forward, but hardly the stuff of science fiction novels or movies where robots take control of mankind! Thankfully we are not yet the subjects of a cruel robot army, but AI has come a long way.

A related term used lately is Machine Learning. There is confusion over whether machine learning is the same as AI, a separate phenomenon, or a "subset" of AI[1]. For our purposes, we refer to machine learning when referring to a computer system that will learn from experience on its own, without specific instructions, and to begin to make decisions or determinations based on that learning. This is an exciting new field, and while it carries risks (what if the computer makes the wrong decision), it has the potential to free up a significant amount of human resources, as we pass decisions and solutions over to computers on our behalf.

Specifically, in the field of trade compliance and supply chain management, machine learning offers tremendous value in several key areas. One whose time has come, is the area of product classification.

Why is Machine Learning needed for product classification?

Before we continue, we should clarify what we mean by "product classification":

In the fields of supply chain management and trade compliance, goods typically need to be classified according to one or more classification schemes. For example, to determine the landed cost of an imported, purchased good, you will need the tariff classification for customs purposes, or Harmonized Tariff Classification (HTS). Similarly, when exporting, you typically need to determine the Export Classification Number (ECN), which will help you determine if a license or permit is required for the export. There are countless other examples of product classification within the supply chain, such as dangerous goods classifications[2], but for now we are referring specifically to HTS and ECN classification.

Focussing on HTS and ECN classification, why is machine learning important, and why are current processes and systems inadequate? To put it simply, these classifications are complex, involving massive lists of classification options, and generally requiring both significant time expenditure, and expertise/experience by the user. On top of the difficulty to classify, most companies have thousands or

---

[1] See: https://www.forbes.com/sites/bernardmarr/2016/12/06/what-is-the-difference-between-artificial-intelligence-and-machine-learning/#7c393182742b

[2] For those using SAP systems, SAP is actively investigation the use of machine learning for dangerous goods classification. See here: https://influence.sap.com/sap/ino/#/campaign/720

more of products requiring a classification. This means one thing to a business: classification is expensive to do right!

Furthermore, beyond the cost of manual classification, this is a serious matter and there is considerable risk in getting classifications wrong. Take HTS for example: the classification chosen dictates, among other things:

- Duty rates
- Special duty rates, such as recent US/China tariff retaliations
- Rule of origin for free trade eligibility
- Other government agency risk targeting, which can lead to delays

In addition to all this, customs will penalize an importer for failing to assign the right HTS, if they discover so under audit. ECN, if anything, is even more serious. Since the ECN dictates whether an export license is needed, incorrect selection of ECN can result in export regulation violations, which carry punishments up to and including prison time.

An automated solution offers the opportunity to be more efficient, and more accurate, than manual classification. Taking into consideration the risks cited above, machine learning has the potential to greatly reduce your cost, and increase your compliance level, if it's done properly.

Application of Machine Learning

How exactly would machine learning work for classification? To help understand this, it's important to understand that a human manually classifying products operates like machine learning. We consider a vast amount of data elements related to the product, and assign a classification based on past feedback and experience.

To help illustrate let's imagine a non HTS or ECN situation, where you are asked to determine if the small furry animal before you is a cat or a dog. Immediately you begin recalling what elements of a cat or dog are distinct: cat's have tails, but so do dogs. Cat's have retractable claws, dogs do not. Cat's have whiskers, dogs do not. Dogs often have long snouts, cats never do. That's probably enough for most situations, let's review:

Our small furry animal has:

- Whiskers
- Retractable claws
- Short snout

I think most people would decide, despite the short snout, that this is probably a cat.

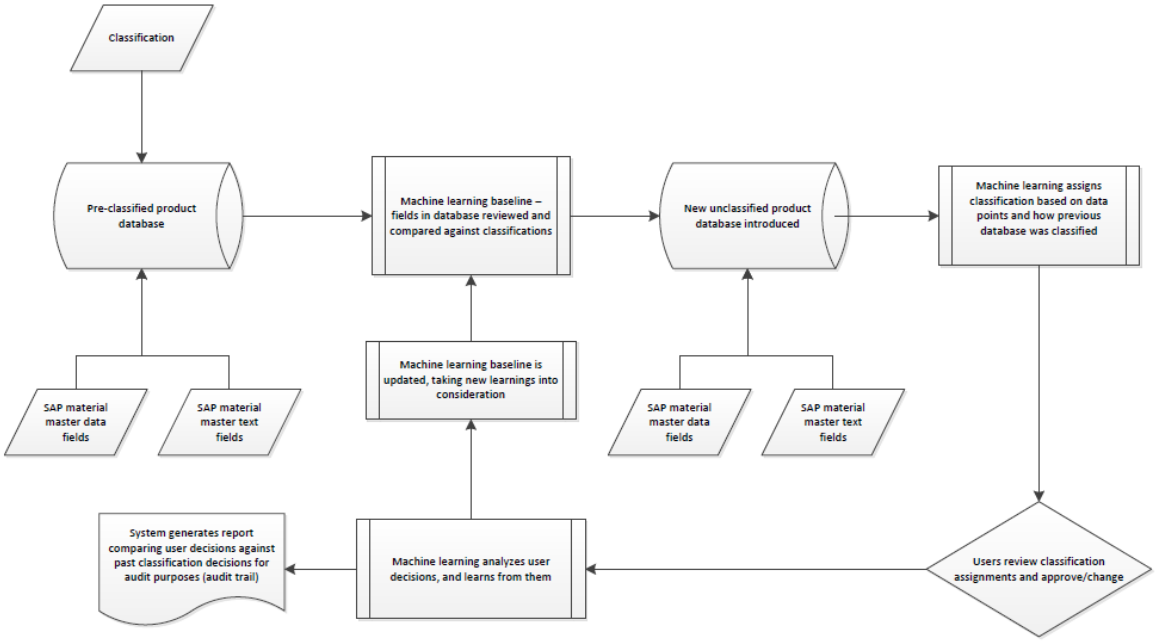
Now, imagine the next thing that happens is: a passerby says: "oh my it's a rare Whatsadoodle Dog! Did you know they are the only dog with retractable claws and whiskers?!"

Well, now you would reconsider your classification, and file in memory that short snouted, retractable clawed and whiskered dogs are called Whatsadoodles! You would also likely start checking other data elements to distinguish this rare dog from a cat (maybe see if it can land on it's feet?).

Machine learning for classification works just like this: it will consider all the applicable data elements, and determine a classification based on how previous classifications were decided, and factoring in feedback. The more data elements and feedback provided, the more powerful the tool can be. Now, let's look at how this process work with a visual aid.

Process flow

Below is a simple representation of how machine learning works with classification:



Please allow a brief walkthrough of what the above flow means. The first step is the feeding of an existing, pre-classified product database (or multiple databases) into the system. This will include as many data elements as possible about each product. Thinking back to our animal example, you want whiskers, claws, snout, etc.…. Using SAP as an example, you want as many elements of the material master as you can get, including but not limited to Product Code, Product Description, Material Group, Product Hierarchy, etc.…. If you can get more data elements beyond what you have in your ERP system, that's even better. For example, if you use SAP GTS, or a TM there may be elements you can draw from there. With Machine learning, the more data the better!

For our process flow, we have divided the data into three broad buckets:

- The product's classification
- Data fields related to the product that are pure "data" i.e. a Material Group, where options are selected from a limited list
- Data fields related to the product that are free-type text based

This will allow the machine learning system to create a baseline, from which future decisions will be made. In truth, this kind of computer logic is not really "Machine Learning" and has existed for a long time. It's just a complex version of logic: "If X = A, then Y". It's what happens next as we continue, that

adds the real value. Following along the process flow, the next step is the introduction of unclassified products. The system analyses the data elements for those products and compares them with it's baseline. Using this combination of data and past data-classification correlation, it assigns a classification to the new products.

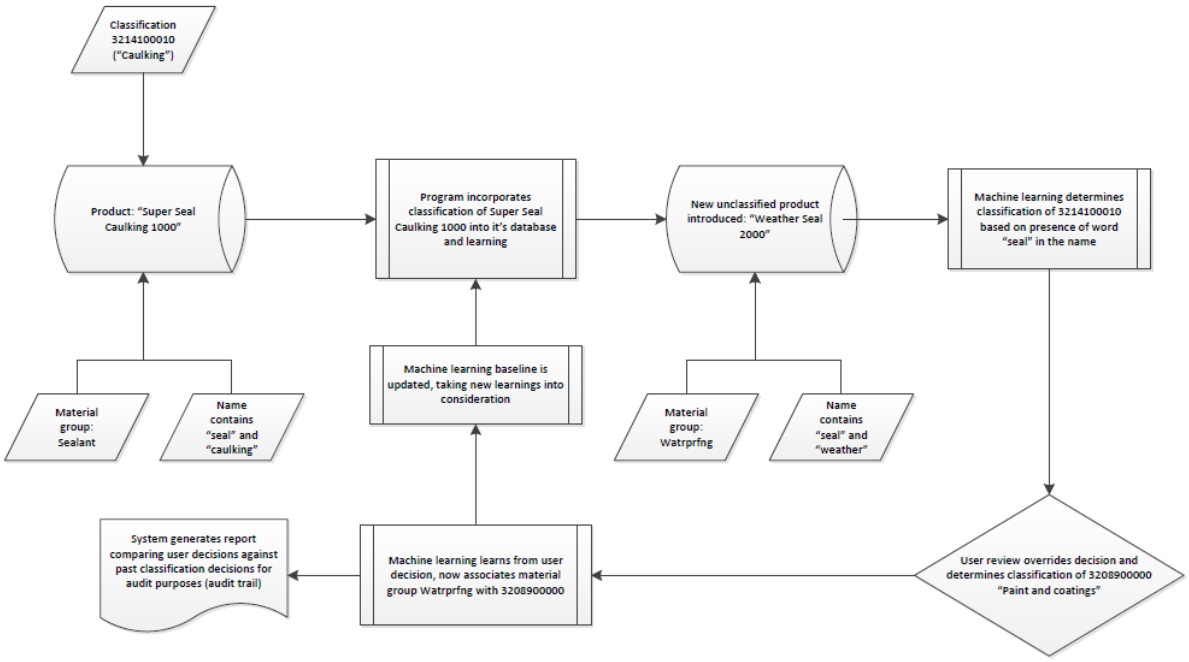
After this comes the step that is essential to machine learning (and human learning for that matter!): feedback. Users will review some or all the classification decisions, and either approve or reject them. Ideally, when rejecting a decision, the user will assign a replacement classification. This is where the "learning" comes in: the system takes the new user feedback into account, and learns from them: what it did right, what it did wrong, and it tries to understand why.

After this feedback process is completed, the system will update its baseline, and come up with a new baseline to judge future classification decisions by. This is a continual and never-ending process, which will result in ever better decisions as time goes on, and more data is introduced. Ultimately the system will perform as well or better than a user can, not to mention faster and more efficiently.

A final step is needed, and that is an audit trail. An attentive reader may have spotted a potential weak point in the process: the user review. If users make bad decisions when reviewing system choices, it can cause the system to make bad decisions in the future. While there are certain system protections available to mitigate this risk (such as consistency checks – ensuring a user's decision doesn't flagrantly oppose previous decisions of a similar nature), an audit trail is essential to ensure the system works right. This report will be periodically reviewed by experts in classification, and act as a check against poor user decisions. As the system gets trusted more and more over time, these audits will need to happen less and less.

Now let's look at an example of this machine learning in action. It is an extremely simple example and takes only a couple data elements into consideration. This is to help illustrate the process, but in actual use there would be many different elements to consider, and the database of reclassified products would likely be quite large as well. Regardless of the scale, however, the same principles will apply:

Existing classification decisions are compared against new unclassified products, resulting in a proposed classification after all relevant data elements are reviewed. User reviews further hone the machine learning, until the system makes choices as good or better than a user would.



The above example uses a simple SAP ERP material master, with a product description and a material group. The system is determining the HTS based on these two fields, and initially makes an incorrect decision. A user corrects the system, giving the system invaluable learning, and future decisions will be more accurate.

Read through the steps of the process, and imagine this happening with millions of products, each containing up to a dozen key data elements, and you can see how refined this process can get. Where the real value comes, is when the system is operated by a 3rd party like Krypt Inc., who has exposure to the databases of multiple customers. This exponentially expands the effectiveness of the system and can offer accurate classifications for new customers immediately.

If you want to learn more about machine learning and classification, talk to your Krypt Inc. contact today, and we will be happy to set up an exploratory meeting with you. We are confident you will see the value in your business, adding both efficiency and compliance.