Canadian Food Inspection Agency | Agence canadienne d'inspection des aliments

Improving Risk Management through Data and Digital Tools

Presented by: Jae-Jin Ryu, Ph.D

Manager, Data Analytics and Modelling

February 28,2019

Canada

Outline

- Canadian Food Inspection Agency (CFIA) at a Glance
- Data-Driven Risk Management Initiatives
  - Canadian Food Safety Information Network
  - Integrated Risk Management
    - Application of Data Science in Regulatory Context
      - Text Analytics, Artificial Intelligence, and Machine Learning

What Does CFIA Do?

- The Canadian Food Inspection Agency (CFIA) is Canada's largest science-based regulatory agency.
- Our business at the CFIA stems from a very broad mandate encompassing food safety, animal health, plant health and international trade.

Food	**Contribute to safeguarding Canada's food supply** • Includes health and safety, nutrition, labelling, recalls
Animal Health	**Protect Canada's animal resource base and Canadians from diseases** • Includes livestock, poultry, animal feeds, and fish and seafood
Plant	**Protect Canada's plant resource base** • Crops, horticulture, nurseries, forest resources and products, greenhouses, seeds, fertilizers, plants with novel traits, invasive alien species
International Trade	**Facilitate market access for Canada's plants, animals and food**

How Does CFIA Manage Risks?

- Some examples include:
  - Design rules and regulations (Safe Food for Canadians, Requirements for humane transport of livestock)
  - Conduct inspections (establishments, shipments, product testing)
  - Issue permits and licenses
  - Communicate with Industry, and Public (food recalls, new outbreaks, guidelines)

Changing Operating Environment

How Is the Agency Responding to Changes?

- Integrate and Collaborate
  - Creation of Information Exchange Network for data on food safety (CFSIN)
- Use Data!
  - Allocate resources where they are needed the most
  - Improve the communication of regulatory materials

Canadian Food Safety Information Network (CFSIN)

- The vision and objectives of the CFSIN were formed following a 2008 outbreak of listeriosis that tragically claimed the lives of 22 Canadians
- The Government of Canada committed to creating a nationally integrated network
- CFSIN is a Federal/Provincial/Territorial (FPT) initiative that will link food safety authorities and food testing laboratories across Canada
- The CFIA is leading the initiative in collaboration with the Public Health Agency of Canada, Health Canada.

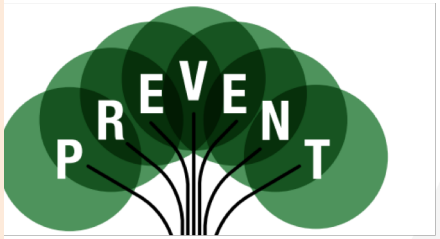
Some of the CFSIN Capabilities

- Contribute to and explore a repository of pan-Canadian laboratory test data.
- Provide a platform to enable CFSIN partners to manage information and actions during events (such as food safety investigations or recalls).
- Carry out basic analysis, visualize data, search for trends and other advanced analysis.
- Share the results of the analysis, via collaboration capabilities, for other CFSIN partners to access.

Key Benefits of the CFSIN

The CFSIN will provide:

Increased ability to detect and prevent food safety incidents



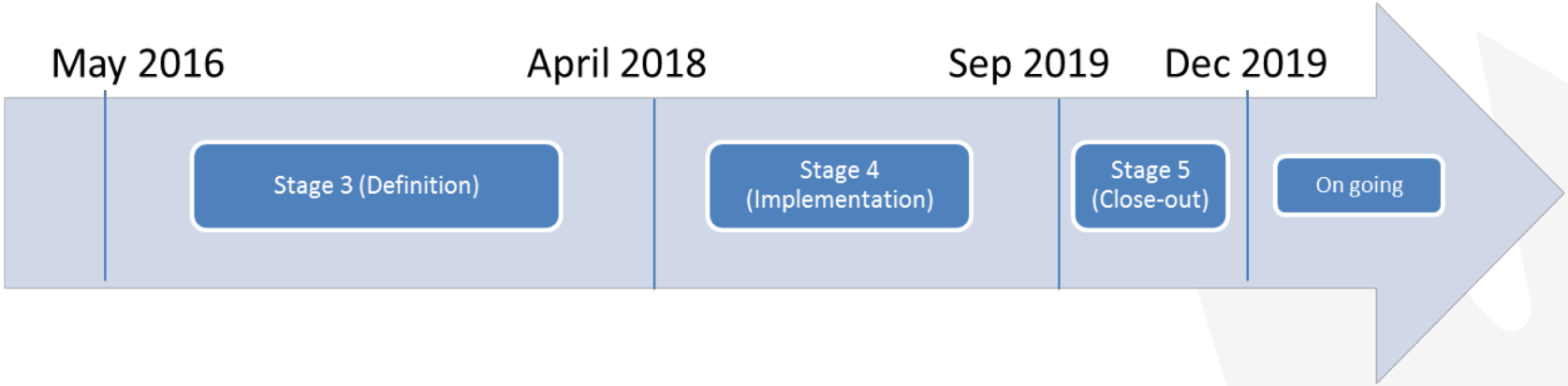
Faster response and resolution of food safety incidents and outbreaks

Responce

Support for international trade

The CFSIN Timeline

- The CFSIN project is in the “Implementation Phase”
- CFSIN will be operational by December 2019



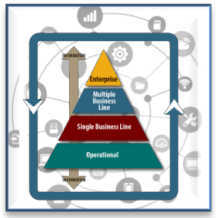
Integrated Risk Management

Using Data Analytics to Guide Our Efforts

BUSINESS INTELLIGENCE

Using Digital Tools & Data Analytics To Be More Effective as Regulators

Data-Driven Risk Management Workflow



Data

Agency holds many sources of structured/unstructured data

Analytics

Digital and analytics tools allow us to gain intelligence and insights

Insights

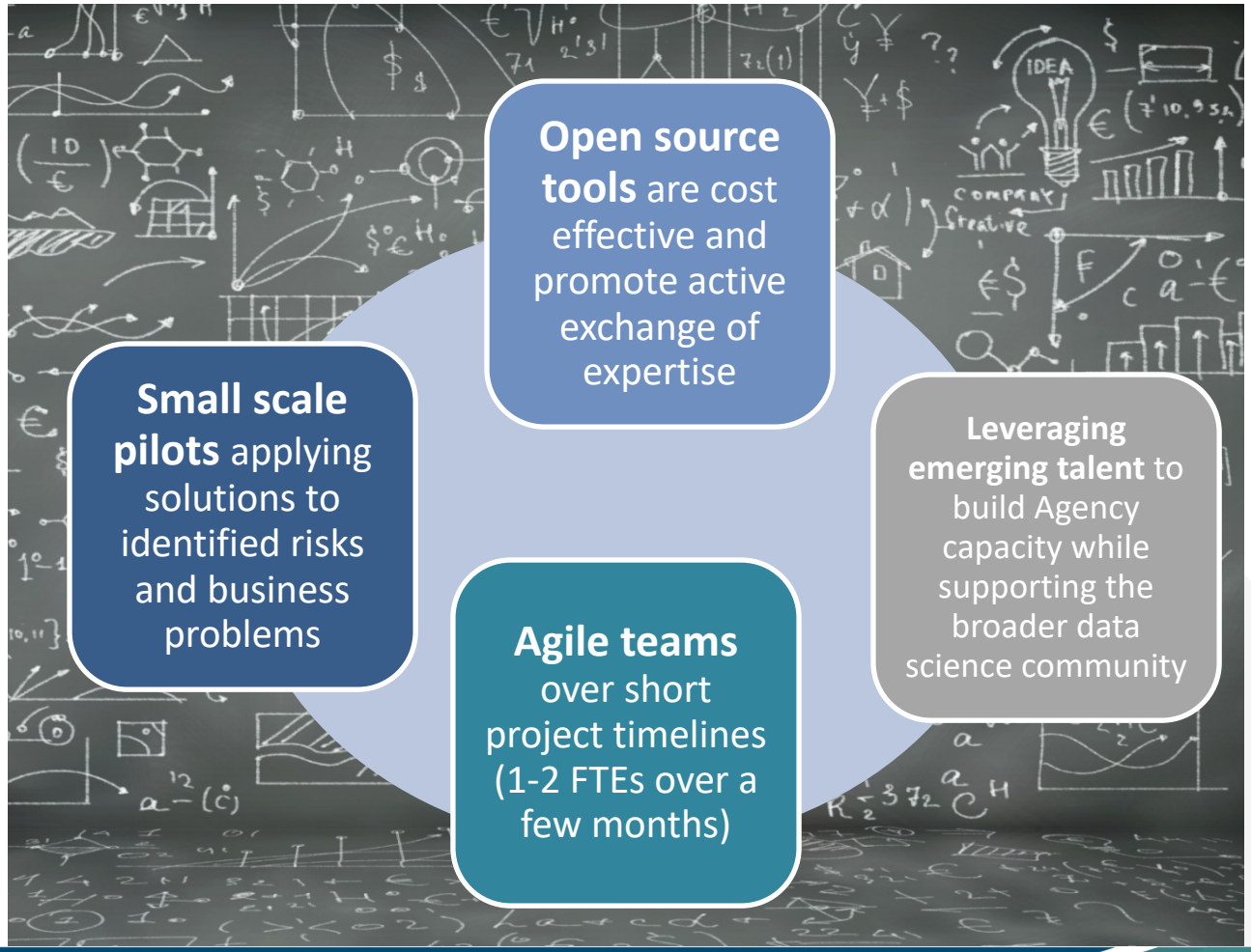
Risk management insights are gained for effective decision making

Application

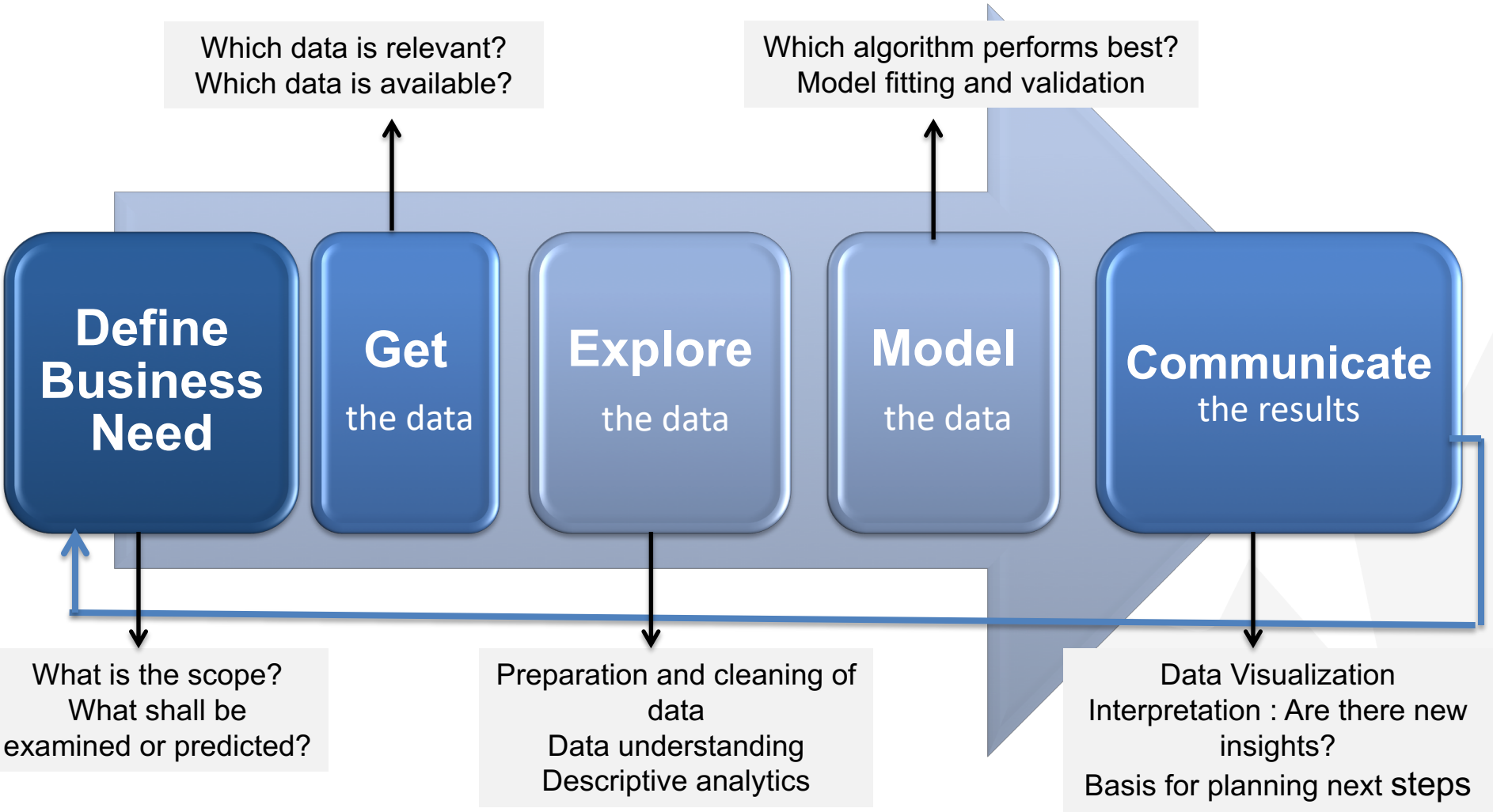
Applications are developed based on insights to target risk mitigation strategies and more proactively address emerging risks

Data Science and the CFIA

Data Analytics and Modelling team



Our Methodology



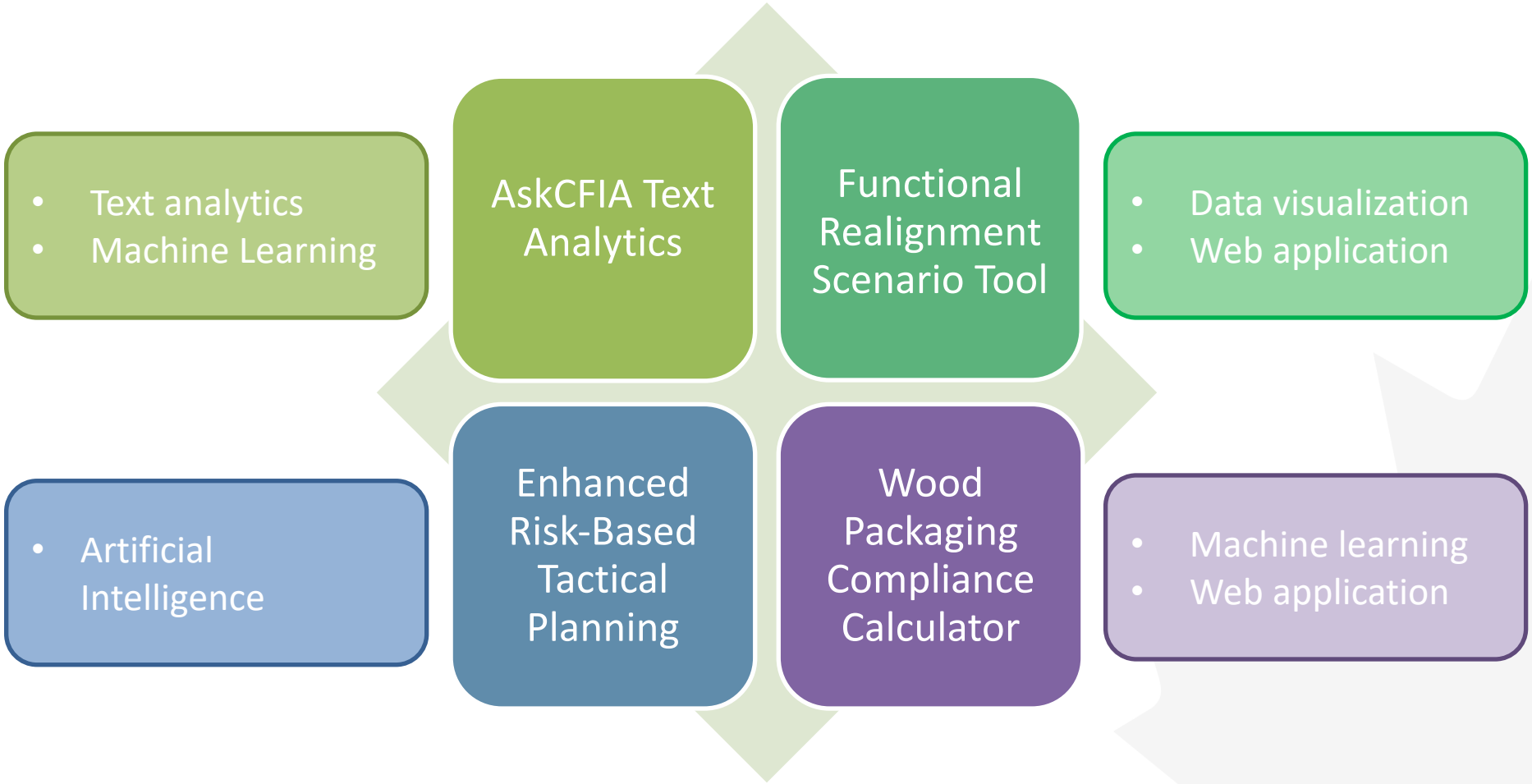
What Kinds of Data Can We Process?

- Database Data
- Data Warehouses
- Transactional Data
- Other Kinds of Data (maps, texts, images audio, etc.)

Which Methods Are We Using?

Data Mining	• The application of particular algorithms for extracting patterns from data
Machine Learning	• A branch of artificial intelligence that tries to define set of approaches to find patterns in data to be able to predict the patterns of future data
Statistics	• A mathematical science that deals with collection, analysis, interpretation or explanation, and presentation of data

Showcase


Text analytics
Machine Learning
AskCFIA Text Analytics
Functional Realignment Scenario Tool
Data visualization
Web application
Artificial Intelligence
Enhanced Risk-Based Tactical Planning
Wood Packaging Compliance Calculator
Machine learning
Web application

AskCFIA Text Analytics

Issue

- Staff review every question received through AskCFIA, triage to appropriate experts, and develop a distinct response. This requires a lot of Agency time and resources.

Business Value

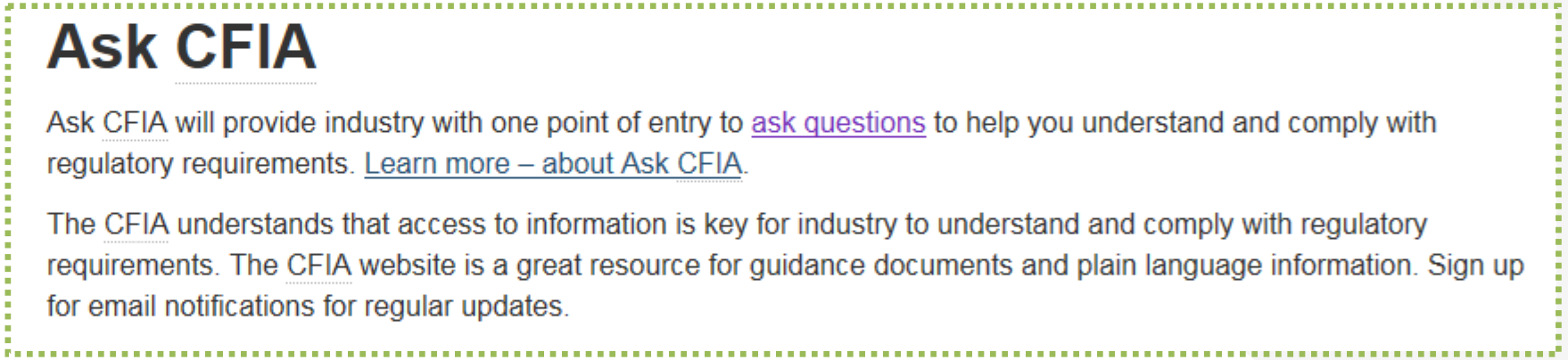
- Less time required to triage, improved client experience
- Automated triaging of questions based on levels of complexity

Our Solution

- Use data science techniques to develop a tool to predict question levels, expedite triaging, and allow for the generation of automated answers where appropriate.
- Cost: 1 student, 1 senior analyst – 3 months

Scalability

- A computer program or AI which converses with clients and provides responses received through AskCFIA or other CFIA-related public communication channels
- Improved user experience: reading level adapted to the user



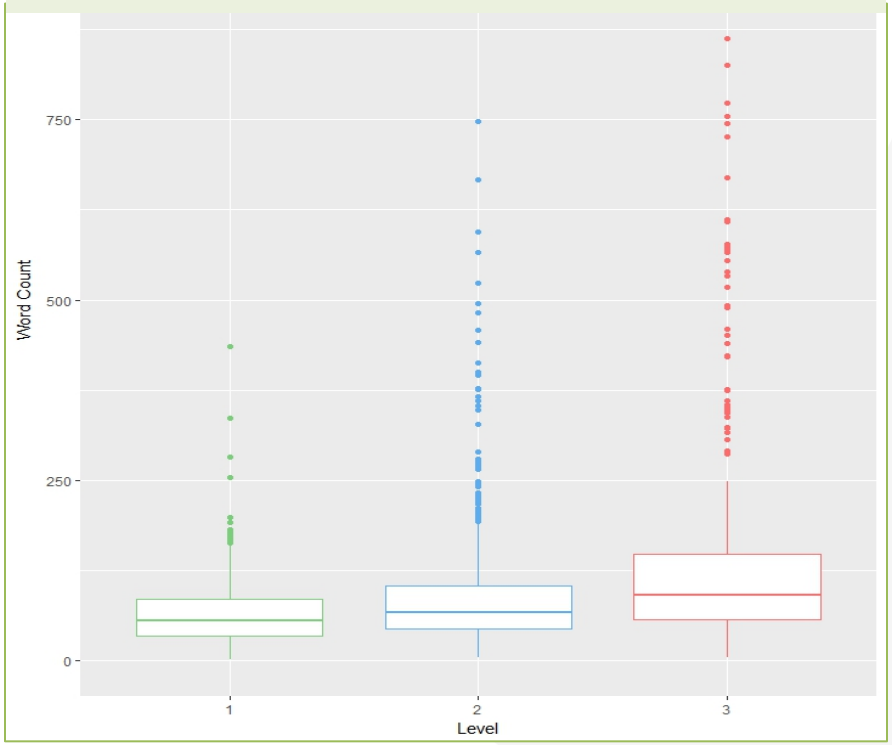
AskCFIA Text Analytics (Cont'd)

Word Cloud

This word cloud shows visually the number of times certain key words were used in AskCFIA questions. The larger words were used the most frequently.

Word Count by Question Level

This chart shows how the word count was closely correlated with the complexity of the question. The most complex questions, were also generally the longest questions in terms of word count.



Enhanced Risk-Based Tactical Planning

Issue

- The current process for annual inspection planning for domestic food establishments is complex and requires significant coordination between a number of different divisions within the Agency to analyze the drivers, identify the risks and the appropriate control measures.

Business Value

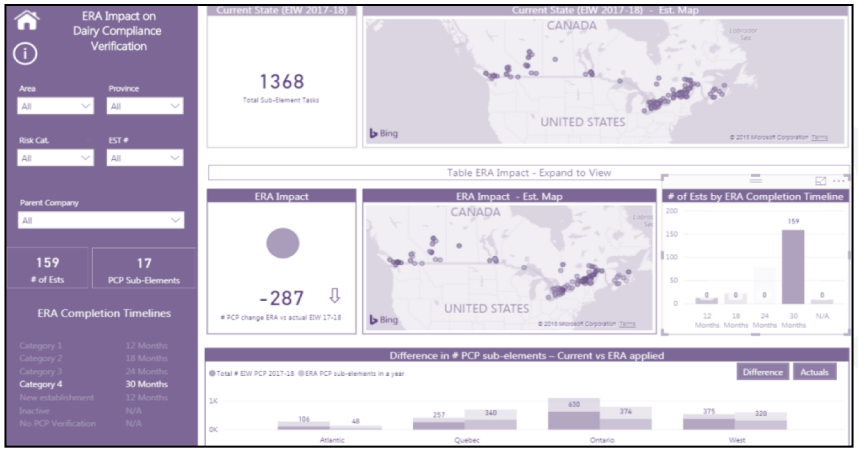
- Reduced activity targeted at low risk establishments to free up capacity to focus on new/higher risk activities.

Our Solution

- Optimize the design, planning and delivery of risk-based establishment inspections based on data and results from various data sets, including compliance history, availability of inspectors, emerging risk information.

Scalability

- Risk information can be used in near real time to enhance program design which informs the priorities for tactical planning and used to consistently guide inspection delivery in the field.



Wood Packaging Compliance Predictor

Issue

- Current shipment selection for compliance verification lacks precision and results in inefficient use of verification resources.

Our Solution

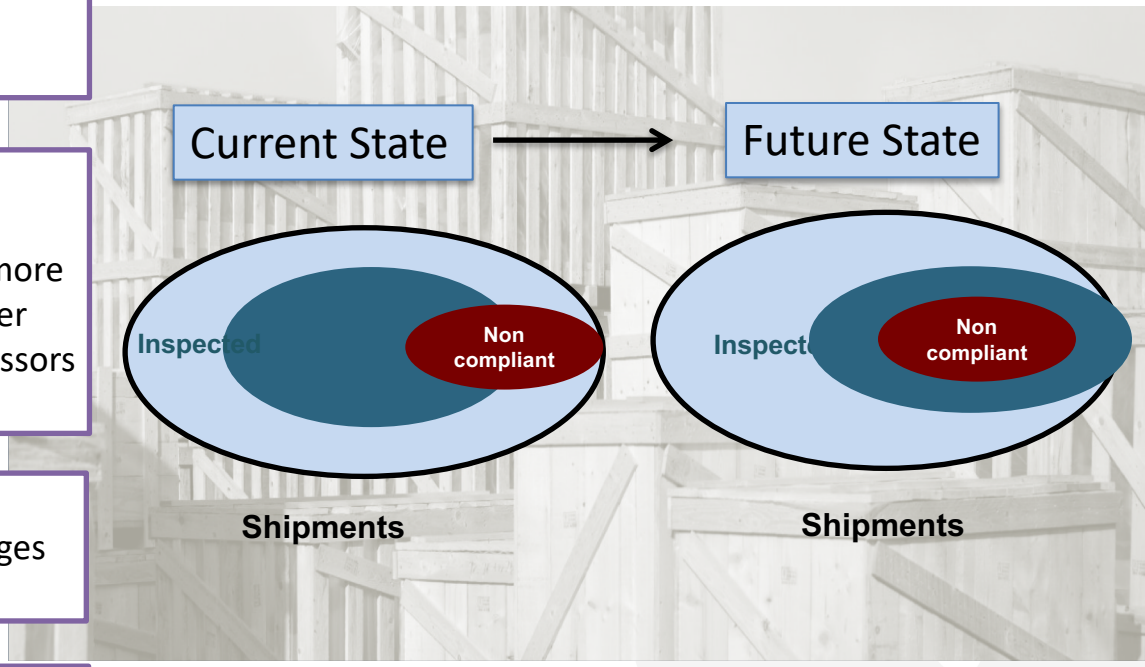
- Use open source tools and machine learning algorithms to consider more risk variables, more accurately predict compliance and build a user friendly interface to assist front-line risk assessors in inspection selection.

Business Value

- Increased detection of non-compliant packages with the same level of inspection efforts

Scalability

- Enhanced targeting of compliance verification activities is broadly applicable e.g. import surveillance, food fraud etc.



Wood Packaging Compliance Predictor (Cont'd)

- A retrospective project
- 85488 observations - between January 1, 2009, and March 31, 2018
- 20 Variables (e.g., country of origin, port of entry, goods description, date of examination, etc).
- Target (Outcome) variable : "Compliance"
- 92% of observations were compliant
  - Uneven or "skewed" distribution of outcome
- Four significant predictor variables: Shipper Country, Port of Entry, Packaging Material, Goods Category

Wood Packaging Compliance Predictor (Cont'd)

Solutions for imbalanced data

Resampling methods

- Random Undersampling
- Random oversampling
- Random Undersampling and Random oversampling
- ROSE
- SMOTE

Methods used

Interpretable Algorithms

- PART decision rules (PART)
- C4.5 decision trees
- Naïve Bayes (NB)
- CART regression trees
- Random Forest

Black-Box Algorithms

- SVM (linear and non-linear)

Wood Packaging Compliance Predictor (Cont'd)

	Current Status (Keyword based)	CART with Undersampling	Non-Linear SVM
Total Inspection Required (%)	**28.45**	**30.5**	**48**
Non-compliant misclassified as compliant (%)	**48.15**	**14**	**2**

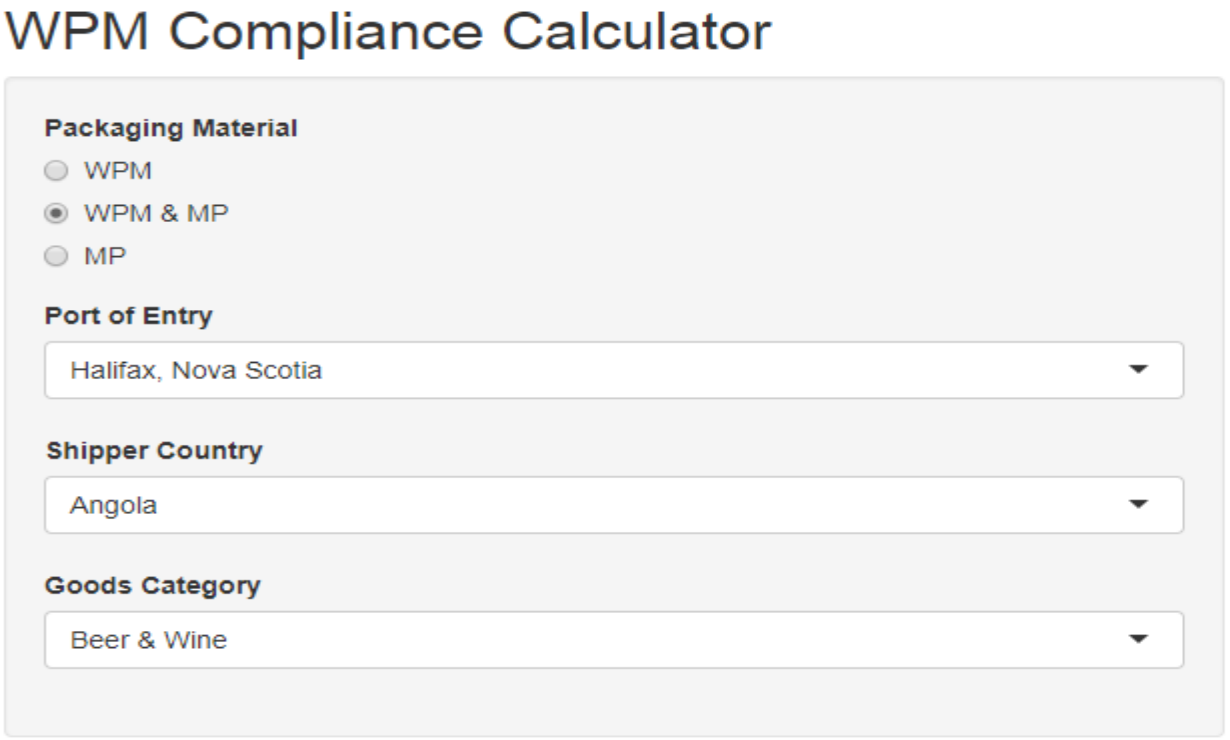
Among all 30 developed models,

- The **CART model with Undersampling** was the best performing model in terms of both F-measure and G-mean.
- The **Non-linear SVM** was the best performing model in terms of Specificity (% non-compliant packages misclassified as compliant)

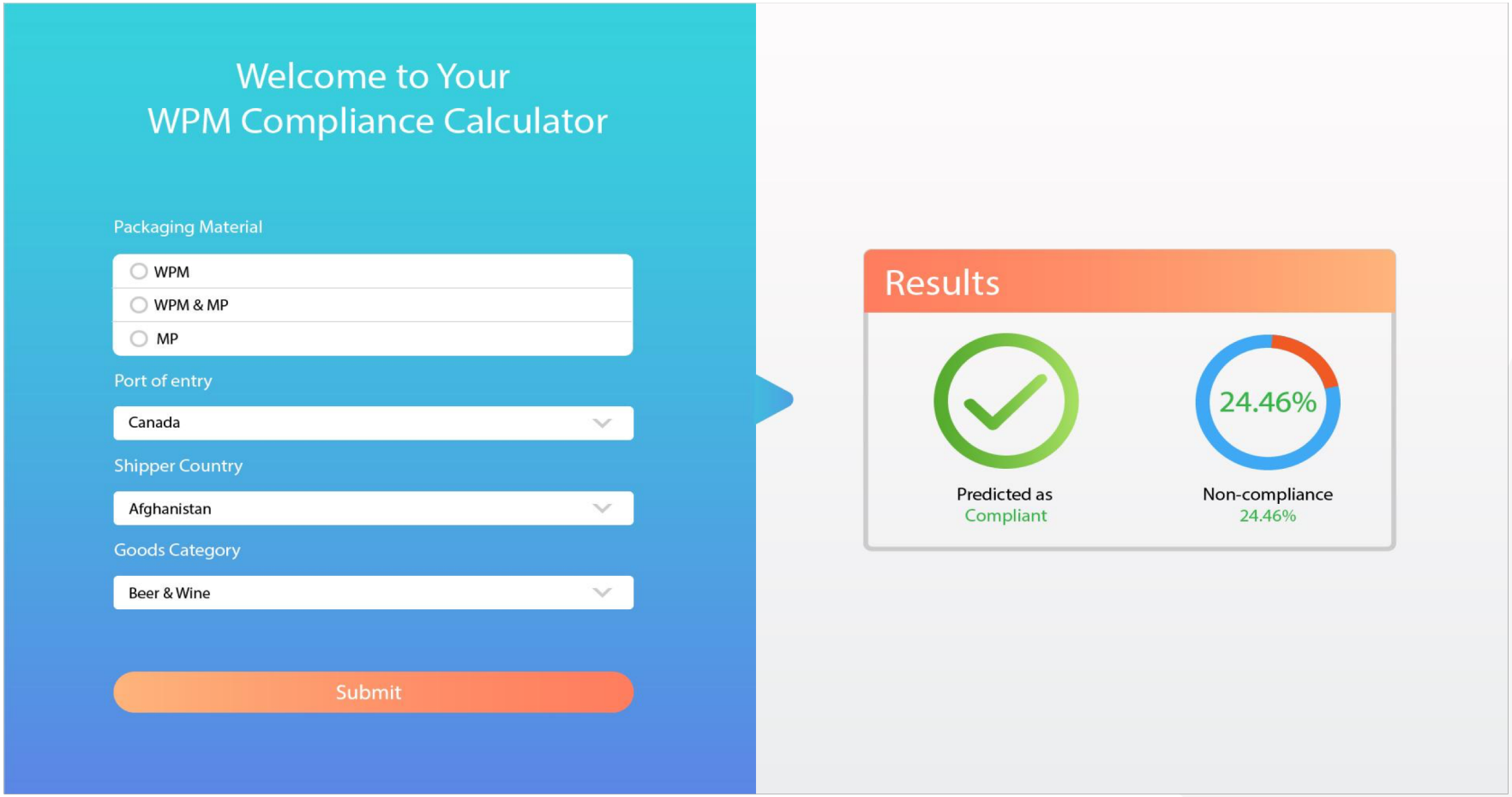
Wood Packaging Compliance Predictor (Cont'd)

A web-based application developed using our model

https://irmmodelling.shinyapps.io/wpm-predictor/



Wood Packaging Compliance Predictor (Cont'd)



Challenges

- The challenge for organizations lies in defining strategies for value generation from the large amount of available data sets
- Applying data science for innovation is not only a technical challenge but impacts the whole organization and its processes
- Data Quality is the key to data science projects
- The right set of people and skills are necessary

Overcoming the challenges requires a collaborative and holistic approach

Data Analytics and Modelling Team

- Github: cfia-data-science
- Analysts: Specialists in data science, emerging technology, economics and food safety
- Co-Op Students: computer science, mathematics, media design, engineering, communication, from Carleton, U Ottawa, Algonquin, Waterloo, U of T

QUESTIONS? SUGGESTIONS?