

# mPowerSmallBiz

W210 Capstone - Final Presentation

Essa, Hoopengardner, Kunjumohamed, Sridhar  
*April 15, 2021*

# Imagine...

- ❖ You are the owner of a small business like Blackrock Fabrication in Dubuque, IA, a company specializing “precision metal fabrication, machining, general manufacturing and assembly tailored to local and regional manufacturers, machine shops, and stamping houses.”
- ❖ Revenue dropped 20% during 2020<sup>1</sup>.
- ❖ You have received \$55k in loans from the \$940B Paycheck Protection Program (PPP), but this assistance is temporary.



How are you going to find new buyers to make up the lost revenue?

<sup>1</sup><https://www.brookings.edu/research/ten-facts-about-covid-19-and-the-u-s-economy/>

# Meanwhile...

- ❖ Global supply chain disruptions.
- ❖ \$400B decline in US imported goods in 2020.
- ❖ 23% decline in the type of commodities Blackrock Fabrication produces (HS Code 731210).



*'I've Never Seen Anything Like This': Chaos Strikes Global Shipping*

The pandemic has disrupted international trade, driving up the cost of shipping goods and adding a fresh challenge to the global economic recovery.



THE WALL STREET JOURNAL

Suez Canal Opens, but Shipping Will Be Snarled for Months

Companies scramble to secure port berths and scarce containers as logistics are upended; a floodgate is going to open'

Feb. 4, 2020

BUSINESS

**Virus Exposes Cracks in Carmakers' Chinese Supply Chains**

Hyundai will idle plants in South Korea, pointing to a shortage of components due to the coronavirus outbreak and hinting at trouble for other manufacturers.

By Jack Ewing, Neal E. Boudette and Geneva Abdul

PRINT EDITION Outbreak Tests Supply Chains Of Automakers | February 5, 2020, Page B1

April 10, 2020

ECONOMY

**Global Trade Sputters, Leaving Too Much Here, Too Little There**

As the pandemic stretches on, consumers and businesses could run short of some products in months to come.

By Ana Swanson



PRINT EDITION Pandemic Scrambles Supply Chains | April 11, 2020, Page A1

<https://www.nytimes.com/2021/03/06/business/global-shipping.html?smid=fb-share>

[https://www.wsj.com/article/suez-canal-opens-but-shipping-will-be-snarled-for-months-11617429947?mod=searchresults\\_pos13&page=4](https://www.wsj.com/article/suez-canal-opens-but-shipping-will-be-snarled-for-months-11617429947?mod=searchresults_pos13&page=4)

<https://techcrunch.com/2021/02/04/elon-musk-pump-100m-into-delivery-logistics-to-ease-supply-concerns/>

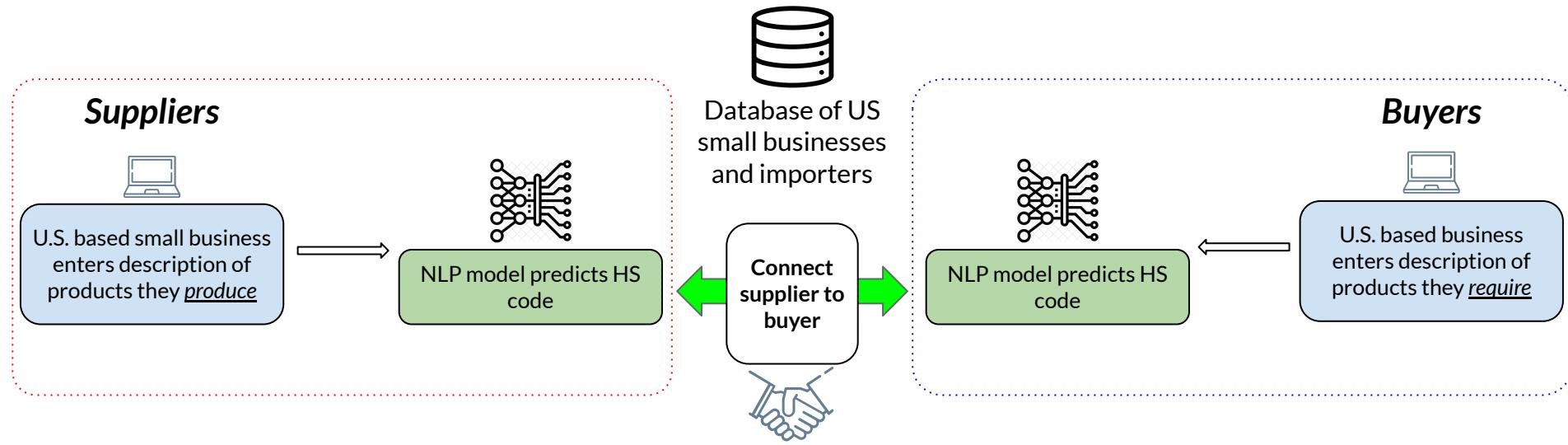
# Nearby...

- ❖ About 300 miles away in Highland, IL is Cooper B-Line, a subsidiary of Eaton, a global power management company.
- ❖ In 2019, Cooper B-Line imported over 120 shipments of the same commodity class that Blackrock Fabrication produces.



Might Cooper B-Line order from Blackrock Fabrication?

# The mPowerSmallBiz Marketplace



# Agenda

- ❖ mPowerSmallBiz Overview (Andy)
- ❖ Data Overview and Approach (Ammara)
- ❖ Modeling Development and Results (Shaji)
- ❖ From Model to Marketplace (Padma)

# Data Overview and Approach

# Training data: Import Bill of Lading Record

```
sampled_df.columns
```

```
Index(['System Identity Id', 'Estimate Arrival Date', 'Actual Arrival Date',  
       'Bill of Lading', 'Master Bill of Lading', 'Bill Type Code',  
       'Carrier SASC Code', 'Vessel Country Code', 'Vessel Code',  
       'Vessel Name', 'Voyage', 'Inbond Type', 'Manifest No',  
       'Mode of Transportation', 'Loading Port', 'Last Vist Foreign Port',  
       'US Clearing District', 'Unloading Port', 'Place of Receipt', 'Country',  
       'Country Sure Level', 'Weight in KG', 'Weight', 'Weight Unit', 'TEU',  
       'Quantity', 'Quantity Unit', 'Measure in CM', 'Measure', 'Measure Unit',  
       'Container Id', 'Container Size', 'Container Type',  
       'Container Desc Code', 'Container Load Status',  
       'Container Type of Service', 'Shipper Name', 'Shipper Address ',  
       'Raw Shipper Name', 'Raw Shipper Addr1', 'Raw Shipper Addr2',  
       'Raw Shipper Addr3', 'Raw Shipper Addr4', 'Raw Shipper Addr Others',  
       'Consignee Name', 'Consignee Address ', 'Raw Consignee Name',  
       'Raw Consignee Addr1', 'Raw Consignee Addr2', 'Raw Consignee Addr3',  
       'Raw Consignee Addr4', 'Raw Consignee Addr Others', 'Notify Party Name',  
       'Notify Party Address ', 'Raw Notify Party Name',  
       'Raw Notify Party Addr1', 'Raw Notify Party Addr2',  
       'Raw Notify Party Addr3', 'Raw Notify Party Addr4',  
       'Raw Notify Party Addr Others', 'Product Desc', 'Marks & Numbers',  
       'HS Code Sure Level', 'CIF', 'Indicator of true supplier',  
       'Indicator of true buyer', 'END', 'HS Code', 'Merged_Description'],  
      dtype='object')
```

Buyer information stored in DB w/ associated HS code

- ❖ 12 Million import Bill of Lading records from 2019 US Customs
- ❖ International Harmonized System (HS) codes [6 digits]: Internationally accepted system of names and numbers used by all countries to classify traded products
- ❖ Need HS code and Product Description to build NLP model
- ❖ Buyer database built by extracting Consignee information and w/ associated HS code

HS Code

6896

871499

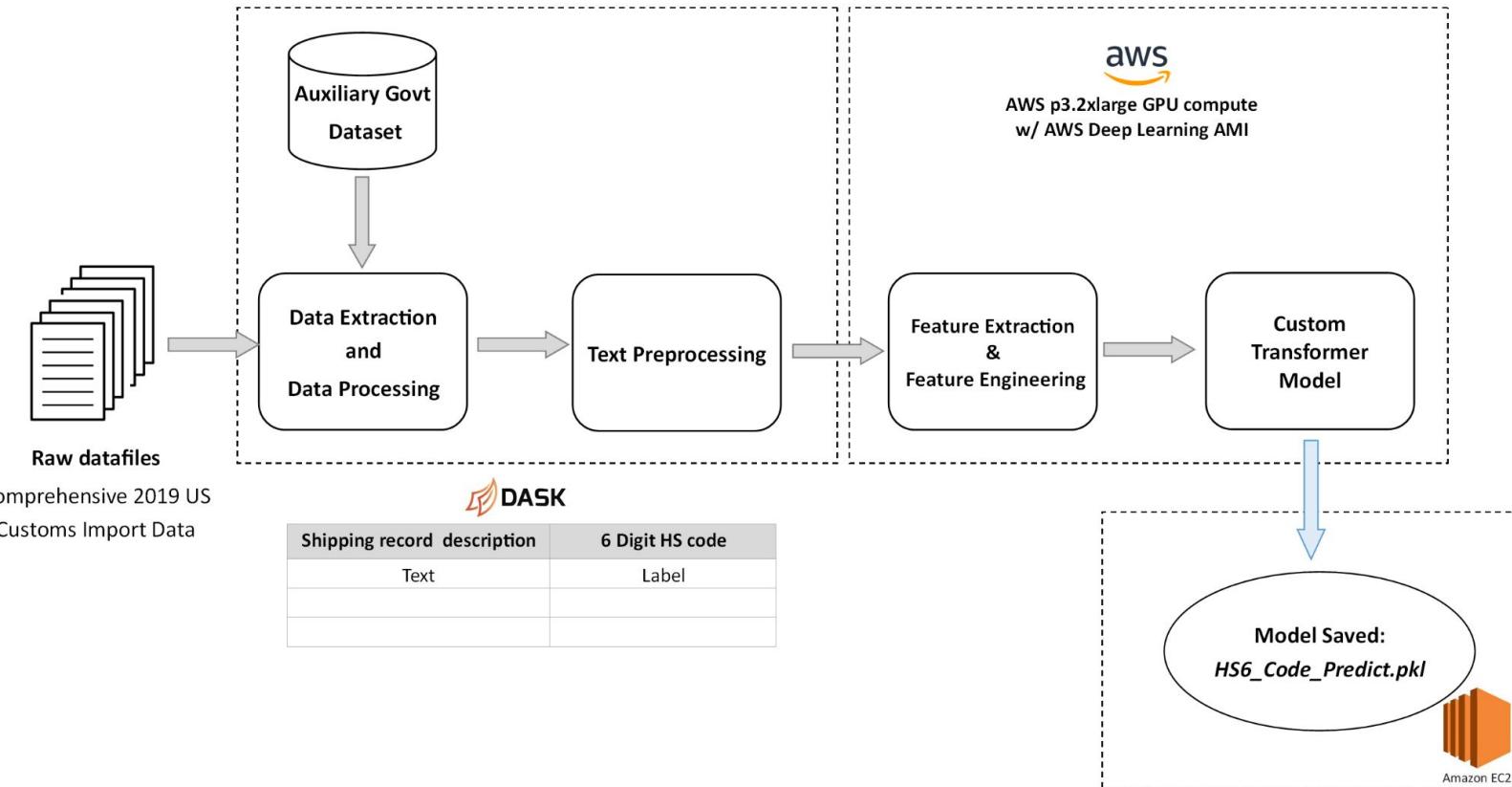
BICYCLE PARTS ALLOY SPACER EXTRUSION FENDER INJECTION MUDGUARD COMPLEX MUDGUARD ALUMINUM FENDER<br/>

Product Desc

# Additional Datasets

- ❖ Supplier database built using PPP (Paycheck Protection Program) data from 2020-2021
  - Loans backed by SBA (Small Business Administration)
- ❖ Rural/Urban Identifier per Zip Code
  - Data source: <https://data.census.gov/>
  - Adding field identifying the ratio between urban and rural population per zip code allows us to prioritize business in rural areas
  - Identifier included for both supplier and buyer

# Modeling Pipeline



# NLP Model Development

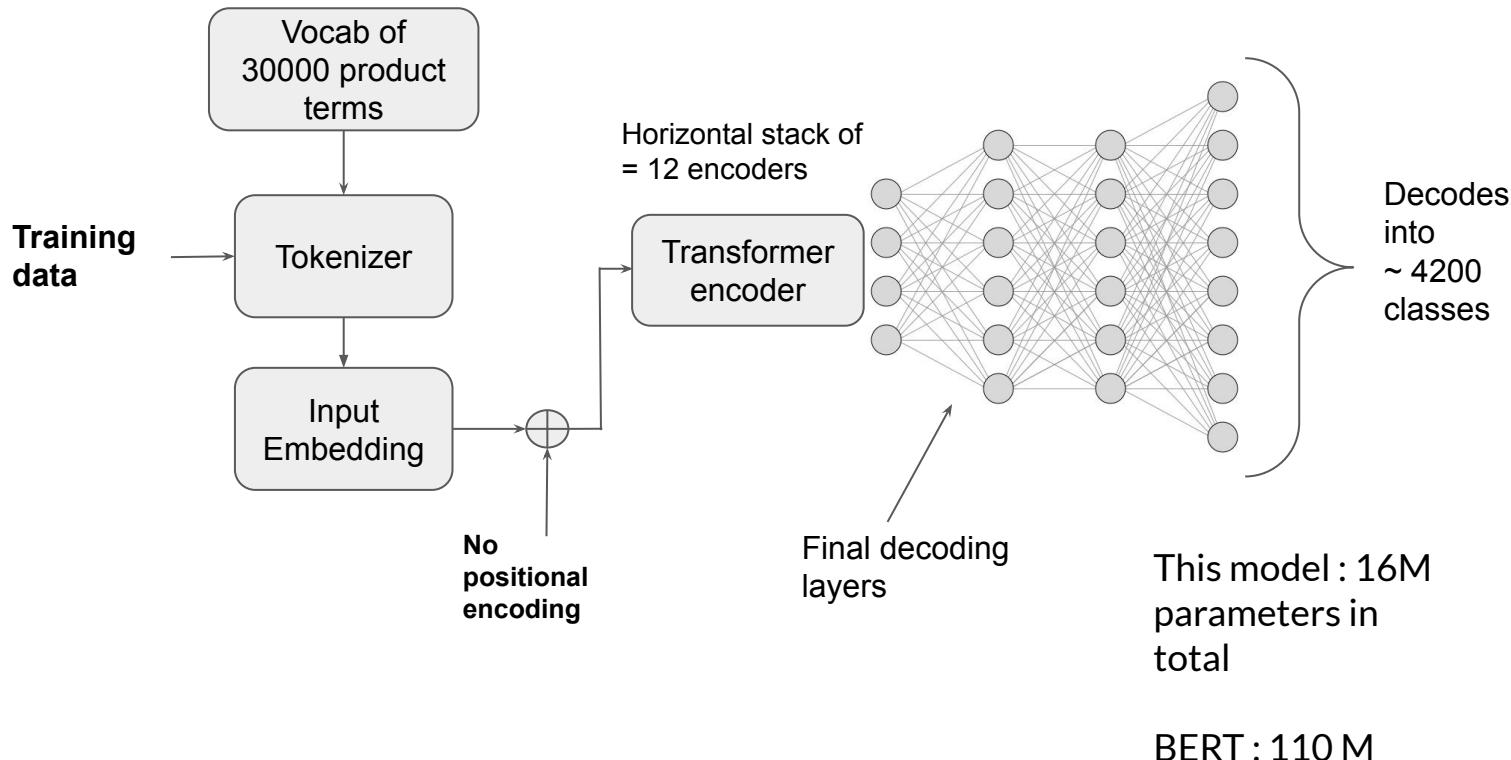
# Modeling approach

Evaluation Metric: Accuracy (sum of top five predictions)

Model	Accuracy			Decision Criteria
	Small dataset* 10 HS Codes (classes) 7K records	Medium dataset* 209 HS Codes (classes) 150K records	Large dataset 4,137 HS Codes (classes) 2M records	
Multinomial Naive Bayes	0.87	0.76	0.48	n/a
Random Forest	0.93	0.85	Failed to run	n/a
XGBoost	0.94	0.91	0.76	<ul style="list-style-type: none"><li>- 500Mb model</li><li>- \$60/training iteration in AWS</li><li>- Cannot scale, convergence issues</li><li>- 17 hour training time</li></ul>
Neural Net with custom Transformer architecture	0.93	0.79	0.81	<ul style="list-style-type: none"><li>- 19Mb model</li><li>- Free, trains in local GPU</li><li>- Scales up for full training data</li><li>- 2 hour training time</li></ul>

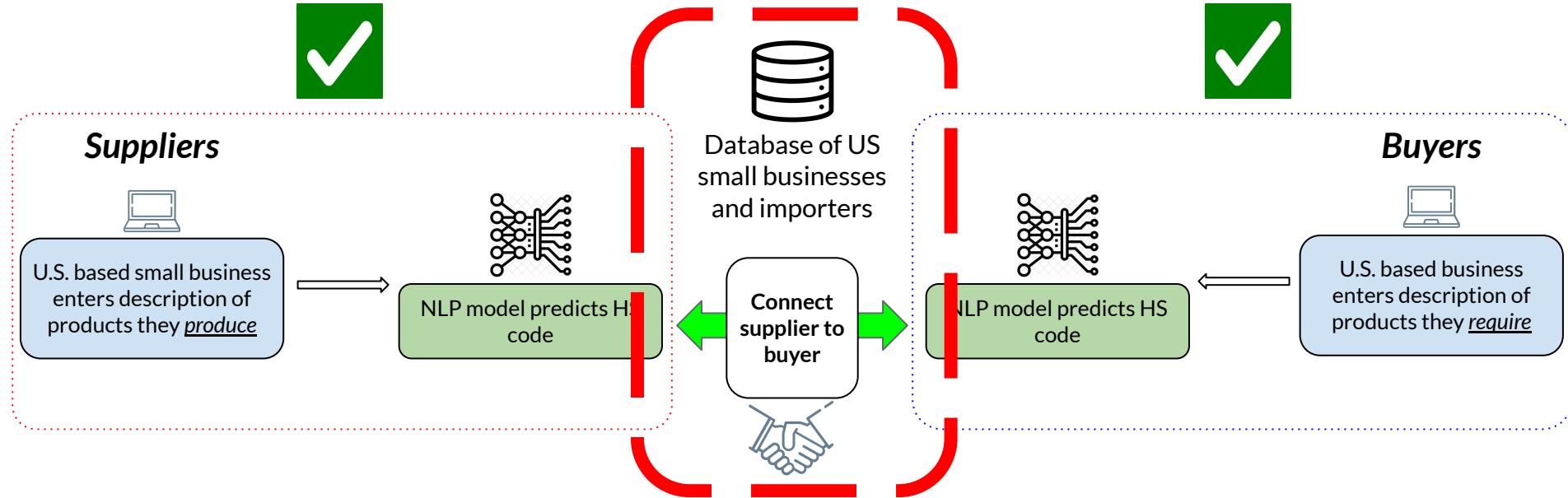
\* Accuracy of top prediction

# Transformer Model Architecture



# From Model to Marketplace

# From model to marketplace



How do we now connect suppliers to buyers?

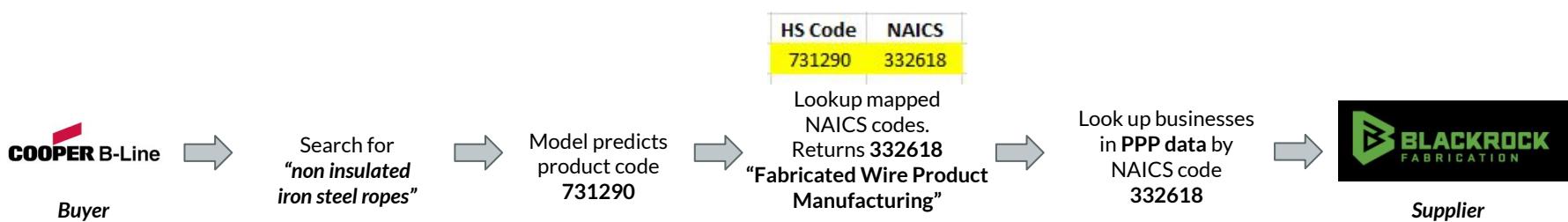
# Marketplace supplemental data

## Buyers

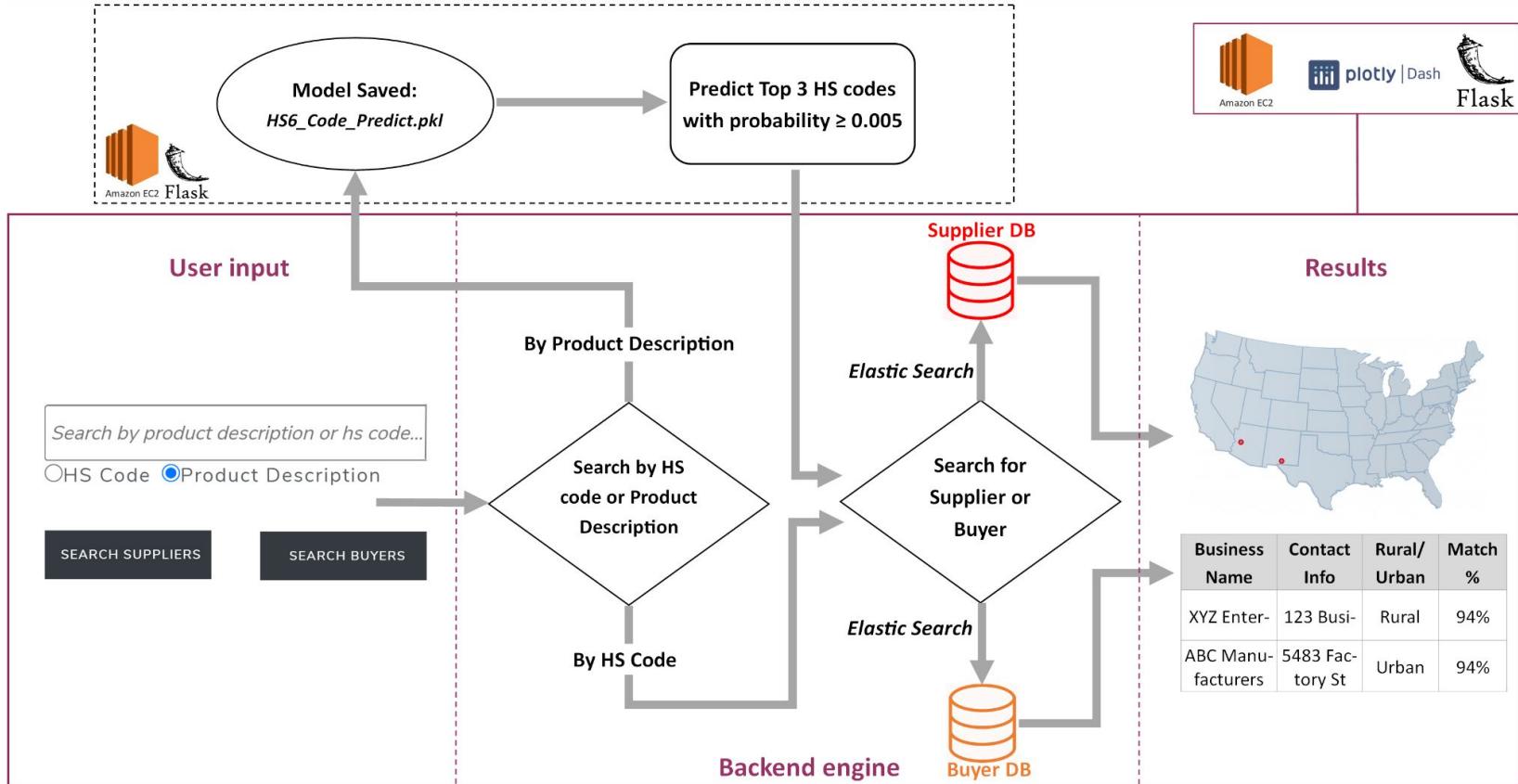
- ❖ U.S. based consignees who purchased imported goods in 2019 i.e.“roller chain”, “flathead screw”
- ❖ Data: US Customs Import data for 2019

## Suppliers

- ❖ U.S. based small/medium business registered and assigned a NAICS code
- ❖ Data: PPP (Paycheck Protection Program) data from 2020-2021
  - Loans backed by SBA (Small Business Administration)



# Website Deployment



# Demo

<http://mpowersmallbiz.com>

# Summary

mPowerSmallBiz:

- ❖ Creates meaningful connections between US businesses
- ❖ Revitalizes local economies
- ❖ Demonstrates a **unique** approach to addressing economic challenges from Covid-19
  - Targeted supply chain development
  - Custom model architecture

# Acknowledgements

- ❖ Indian School of Business - primary research stakeholder, provider of 2019 US import customs data.
- ❖ Reid Hutchins - Managing Director, Advantage Capital, an investment firm that has invested \$3.6B in 780+ companies since its founding in 1992.
- ❖ Taylor Bergmann - Demand Planner, MilliporeSigma (subsidiary of Merck).
- ❖ Brian Dunch - Principal (Partner), PwC Advisory.
- ❖ Martin Cresmer - General Manager at Pier Maintenance Industrial, LLC Los Angeles County, California, United States.
- ❖ Prof. Richard H. Steinberg - Professor UCLA, International Trade Law (subject matter expert).



# Questions?

# Team Contributions

# Team Contribution: Ammara

- Supplemental data procurement
- Exploratory Data Analysis
- Dataset processing/data extraction
- Initial ML model exploration
- Architecture visualizations
- Project management

# Team Contribution: Andy

- Exploratory Data Analysis
- Initial ML model exploration
- Chapter-level model development
- Website design
- Presentation development
- Project management

# Team Contribution: Shaji

- Infrastructure setup and support
- Exploratory Data Analysis
- Initial ML model exploration
- HS Code-level model development
- Visualization development
- Initial website demo

# Team Contribution: Padma

- Data procurement
- Data preparation and pipeline
- Exploratory Data Analysis
- Initial ML model exploration
- Website development
- Primary stakeholder engagement

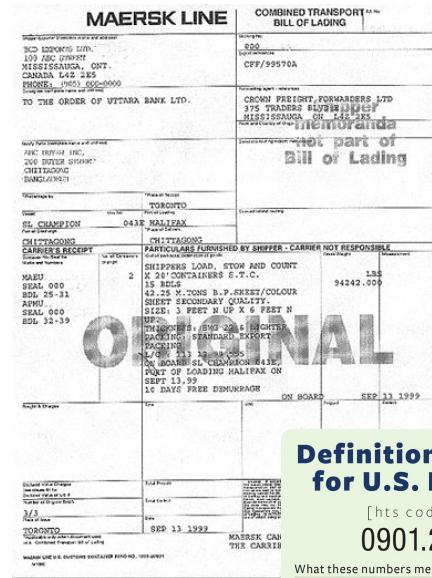
# Appendix

# References

- ❖ Luppes, J. (2019). *Classifying short text for the harmonized system with convolutional neural networks.* (Master's thesis, Radboud University,Netherlands)
- ❖ Altaheri, Fatma & Shaalan, Khaled. (2020). *Exploring Machine Learning Models to Predict Harmonized System Code.* 10.1007/978-3-030-44322-1\_22.
- ❖ *ManifestDB - Exports & Imports World Trading Data*, [www.manifestdb.com](http://www.manifestdb.com).
- ❖ *World Integrated Trade Solution (WITS)*, <https://wits.worldbank.org/> .
- ❖ *International Trade Administration | Trade.gov*, [www.trade.gov](http://www.trade.gov)/.
- ❖ U.S. International Trade Commission - USITC. *Harmonized Tariff Schedule Search*, [hts.usitc.gov/](http://hts.usitc.gov/).
- ❖ “*Explore Census Data*, [data.census.gov/](http://data.census.gov/).

# What are HTS codes?

- ❖ International Harmonized System (HS) codes [6 digits]: Internationally accepted system of names and numbers used by all countries to classify traded products
- ❖ U.S. Harmonized Tariff Schedule (HTS) [7-10 digits] : comprises of 6 digit HS code + additional codes to further classify an item
- ❖ Proper product classifications crucial for international trade
  - Tariffs and duties are imposed on imports/exports based on the HS/HTS classification code
  - Conduct market research and obtain trade statistics
  - Monitoring controlled goods
- ❖ Missing or incorrect HTS codes can lead to delays and penalties



## Definition & Example for U.S. HTS Codes

[hts code example]  
**0901.21.0010**

What these numbers mean

**09**  
Chapter

Coffee, Tea, Mate And Spices

**0901**  
Heading

Coffee, Whether Or Not Roasted Or Decaffeinated; Coffee Husks And Skins; Coffee Substitutes Containing Coffee

**0901.21**  
Sub Heading (HS Code)

Coffee, Roasted, Not Decaffeinated

**0901.21.00**  
Subheading (Determines Duty)

No Distinction

**0901.21.0010**  
Statistical Suffix

Coffee, Roasted, Not Decaffeinated, Certified Organic

DESCARTES  
Datamyne

# 2020 Small Business Revenue Change

FIGURE 1.

Change in Small Business Revenue for Selected Industries Relative to January 2020,  
January–August 2020



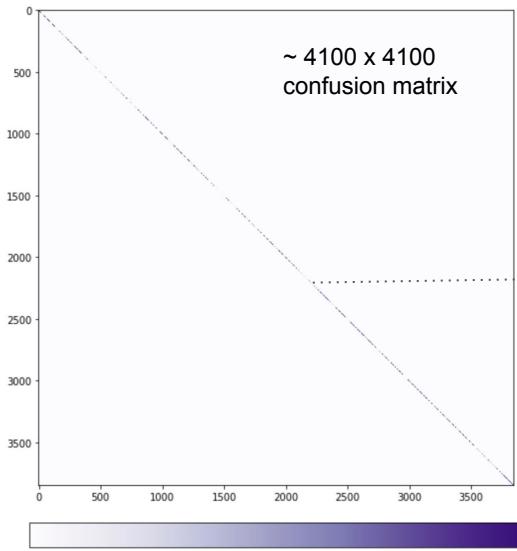
Source: Chetty et al. 2020.

Note: Data are for January 10, 2020 to August 9, 2020. Raw data are collected from Womply, a firm-based panel data set of small business revenue. "Revenue" is defined as the sum of all credits (generally purchases) minus debits. Data are seasonally adjusted, calculated as a seven-day moving average, and indexed to January 4–31, 2020. Chetty et al. (2020) analyze Womply data at the two-digit industry level and then aggregate using NAICS supersector codes. "All industries" includes both the shown supersectors as well as the other supersectors (e.g., manufacturing, financial services, etc.). For more on supersectors, see BLS 2019.

THE  
**HAMILTON**  
PROJECT  
BROOKINGS

<https://www.brookings.edu/research/ten-facts-about-covid-19-and-the-u-s-economy/>

# For Appendix : Model evaluation - confusion matrix



**871499 (Actual)** : Bicycle twist grips and levers, trigger and twist grip controls for three speed hubs bicycle handlebar stems

**851210 (Predicted)** : Lighting or visual signaling equipment used on bicycles

	Predicted value																			Actual value		
	HTS 401693	HTS 841420	HTS 842951	HTS 848330	HTS 851210	HTS 870829	HTS 871200	HTS 871410	HTS 871420	HTS 871491	HTS 871492	HTS 871493	HTS 871494	HTS 871495	HTS 871496	HTS 871499	HTS 901600	HTS 902690	HTS 902890	HTS 902990		
HTS 401693	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
HTS 841420	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
HTS 842951	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
HTS 848330	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
HTS 851210	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
HTS 870829	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
<b>HTS 871200</b>	0	1	1	0	0	0	<b>174</b>	0	1	1	1	0	0	0	0	1	0	0	0	0	0	
<b>HTS 871410</b>	0	0	0	0	0	0	2	0	<b>162</b>	0	0	0	0	0	0	1	0	1	0	1	0	
<b>HTS 871420</b>	1	0	0	0	0	0	0	0	<b>186</b>	0	0	0	0	0	0	<b>3</b>	<b>2</b>	0	0	0	0	
<b>HTS 871491</b>	0	0	0	0	0	0	0	2	0	0	<b>56</b>	0	0	0	0	1	0	0	0	0	0	
<b>HTS 871492</b>	0	0	<b>2</b>	0	0	0	0	0	0	0	<b>192</b>	1	0	0	1	1	0	0	0	0	0	
<b>HTS 871493</b>	1	0	0	0	0	0	0	0	0	0	0	<b>177</b>	0	0	0	0	0	0	0	0	0	0
<b>HTS 871494</b>	0	0	0	0	1	0	0	0	0	1	0	0	<b>136</b>	0	0	0	0	0	0	0	1	0
<b>HTS 871495</b>	0	0	0	0	0	0	0	1	0	0	0	0	0	<b>12</b>	0	<b>2</b>	0	0	0	0	0	0
<b>HTS 871496</b>	1	1	0	<b>2</b>	0	0	1	0	0	0	0	0	0	0	<b>67</b>	0	0	0	0	0	1	
<b>HTS 871499</b>	0	<b>3</b>	0	0	<b>3</b>	0	0	0	0	1	0	0	0	0	0	<b>153</b>	0	0	<b>2</b>	0	0	
HTS 901600	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
HTS 902690	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
HTS 902890	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	
HTS 902990	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	