Multiclass Classification Using Deep Learning

Nancy Hamdan

Table of Contents

Executive Summary

This report presents findings from exploring the Military Equipment for Local Law Enforcement dataset that is provided from The Defense Logistics Agency (DLA) in the United States. Excess military equipment and devices from the Department of Defense in the US are handled by the DLA which transfers this excess equipment to federal, state, and local law enforcement agencies in the US as part of the LESO program (Law Enforcement Support Office program). The dataset contains 130958 records representing the transfers that happened between 01/01/1980 and 30/09/2021.

All military items that are transferred must be demilitarized to some degree; the degree of demilitarization required is represented by the DEMIL Code attribute in the dataset. Each DEMIL Code has a corresponding DEMIL IC representing the integrity of the DEMIL Code registered for the transfer. This report details the process that was followed in building a high performing deep learning predictor of the DEMIL Code and in deriving key insights about the data.

Upon exploring the dataset, the following key insights were found:

- 71% of the transfers were of the class 'D'.
- 85% of the transfers had a DEMIL IC of 1. DEMIL IC of 1 is the highest level of integrity and thus most DEMIL Codes possessed high integrity.
- Three out of the five topmost transferred items are weapons (rifles and pistols) indicating increased militarization of local police enforcements.
- Over 63% of all non-reviewed DEMIL Codes (those that had blank DEMIL IC) are DEMIL Code 'A'. As DEMIL Code 'A' is for low-risk items that are released out of DoD control, these items' DEMIL Codes' possessing the least integrity poses a risk.
- Items with non-reviewed DEMIL Codes are the second highest in terms of quantity and acquisition value transferred over the years. This poses a risk too.
- The data indicates that structural changes have happened in September 2009 and June 2020 that caused the quantities of items transferred to dramatically spike. Each structural change caused the quantities of items transferred to have an increasing trend in the subsequent time periods of the changes.

The best predictor model for predicting the DEMIL Code is a deep neural network with 6 hidden layers that uses both one hot encoding of the features as well as embeddings of them to predict the DEMIL Code. This DNN has an accuracy of 98% with high precision and good recall for all classes of the DEMIL Code. The tuned DNN has high recall scores for classes including highly underrepresented ones.

Data Understanding and Exploration

General EDA

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 130958 entries, 0 to 398
Data columns (total 11 columns):
 #   Column             Non-Null Count   Dtype
---  ------             --------------   -----
 0   State              130958 non-null  object
 1   Agency Name        130958 non-null  object
 2   NSN                130958 non-null  object
 3   Item Name          130958 non-null  object
 4   Quantity           130958 non-null  int64
 5   UI                 130958 non-null  object
 6   Acquisition Value  130958 non-null  float64
 7   DEMIL Code         130958 non-null  object
 8   DEMIL IC           121048 non-null  float64
 9   Ship Date          130958 non-null  datetime64[ns]
 10  Station Type       130958 non-null  object
dtypes: datetime64[ns](1), float64(2), int64(1), object(7)
memory usage: 12.0+ MB
```

- DEMIL IC is the only column with missing values. However, using the information given about the dataset, we know that a blank DEMIL IC corresponds to a non-reviewed DEMIL Code, thus no integrity was recorded for that DEMIL Code.
- For the rest of the analyses, a blank DEMIL IC was replaced by -1.

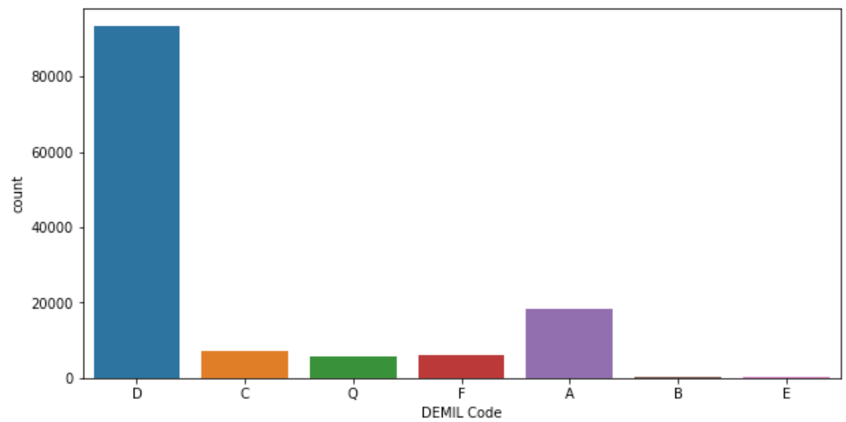
	Quantity	Acquisition Value	Ship Date
count	130958.000000	1.309580e+05	130958
mean	4.180233	1.316023e+04	2011-09-05 01:02:48.702233088
min	0.000000	0.000000e+00	1980-01-01 09:07:07.000032
25%	1.000000	1.380000e+02	2006-05-02 00:00:00
50%	1.000000	4.990000e+02	2012-04-18 00:00:00
75%	1.000000	7.490000e+02	2017-04-17 00:00:00
max	5000.000000	2.200000e+07	2021-09-30 00:00:00
std	31.187727	1.371933e+05	NaN

- Quantity column appears to have outliers, as 75% of the dataset has quantity of 1 but the max quantity is 5000.
- Acquisition Value has outliers too as 75% of the dataset has acquisition value equal to or less than 749 dollars while the maximum is 22,000,000 dollars.
- Ship dates of transfers range from the year 1980 to 2021.

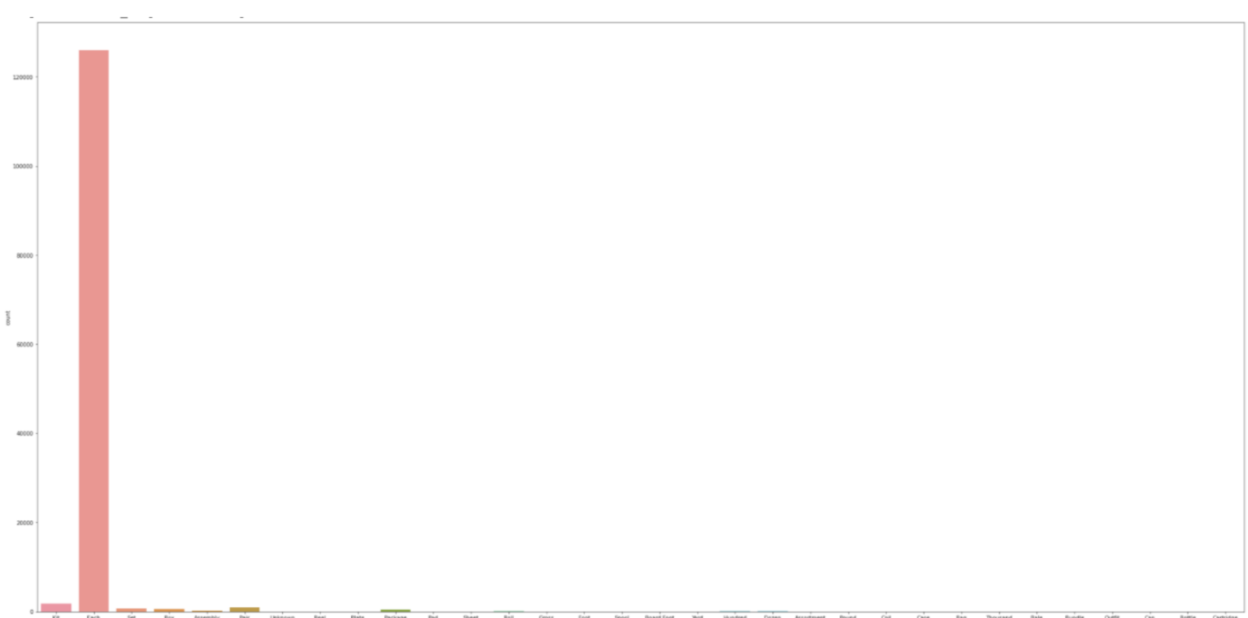
	State	Agency Name	NSN	Item Name	UI	DEMIL Code	DEMIL IC	Station Type
count	130958	130958	130958	130958	130958	130958	130958.0	130958
unique	53	5632	10295	4797	32	7	8.0	1
top	CA	TEHAMA CSO	1005-00-073-9421	RIFLE,5.56 MILLIMETER	Each	D	1.0	State
freq	9489	1402	31543	46843	125920	93478	112444.0	130958

- Most of the dataset has UI=Each. 125920 out of 130958 records are UI=Each.
- Station type is the same for all records, it is "State".
- Over 35% of transfers were of RIFLE,5.56 MILLIMETER items.
- Most of the transfers had a DEMIL code of 'D'.
- Most of the transfers had a DEMIL IC of 1.

<matplotlib.axes._subplots.AxesSubplot at 0x7ff4b4e52350>

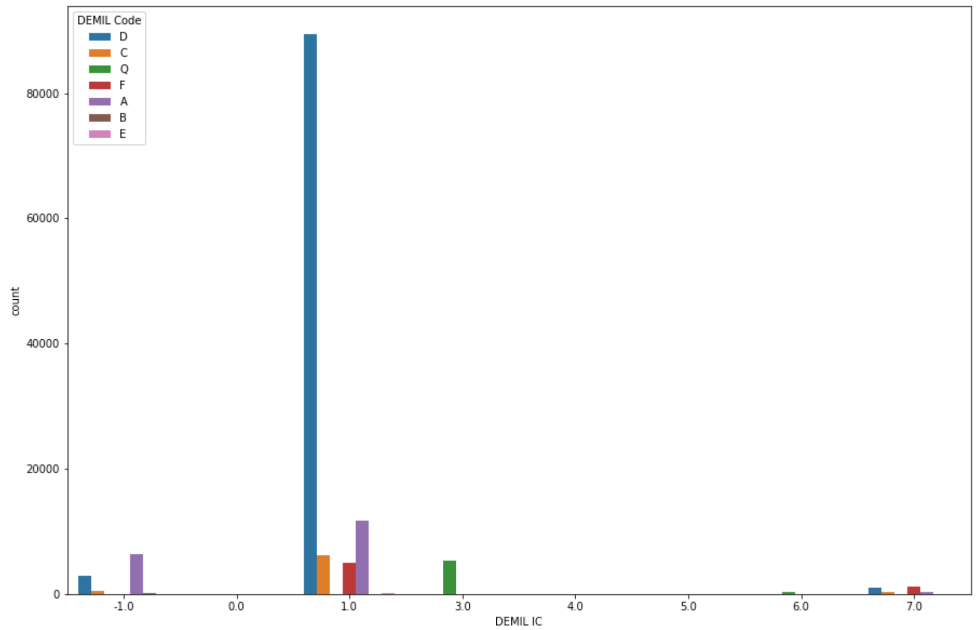


- Class distributions are highly imbalanced with class ‘D’ making more than 71% of the transfers.
- Classes ‘B’ and ‘E’ are especially underrepresented.

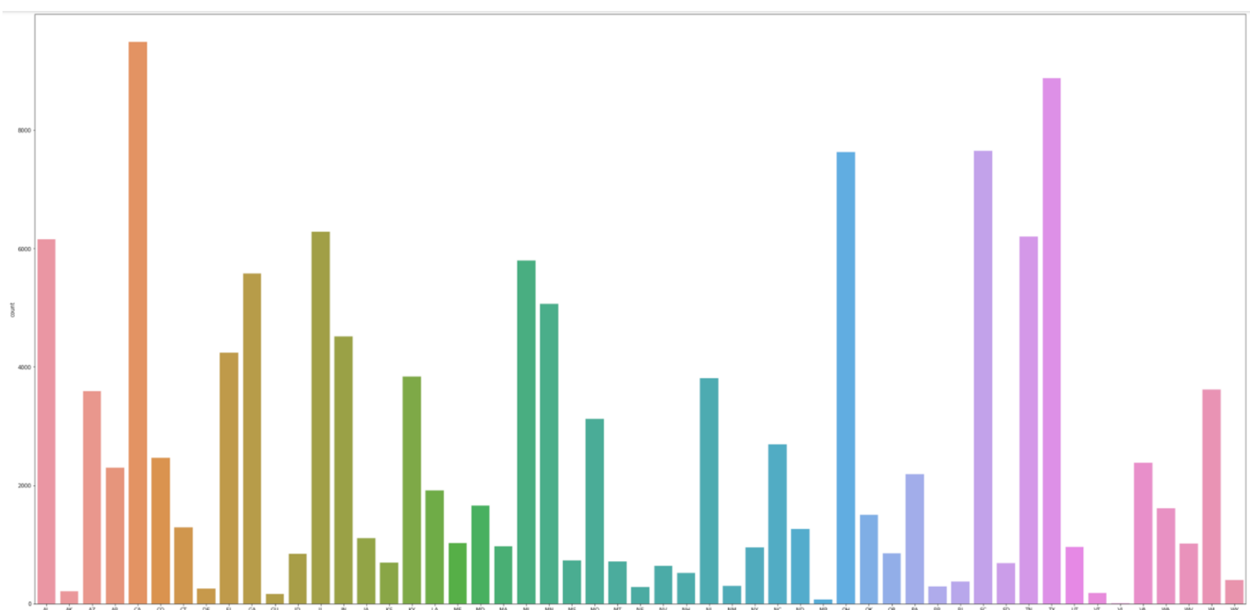


- Over 96% of the data are transfers with UI as each.

<matplotlib.axes._subplots.AxesSubplot at 0x7ff5417f1890>

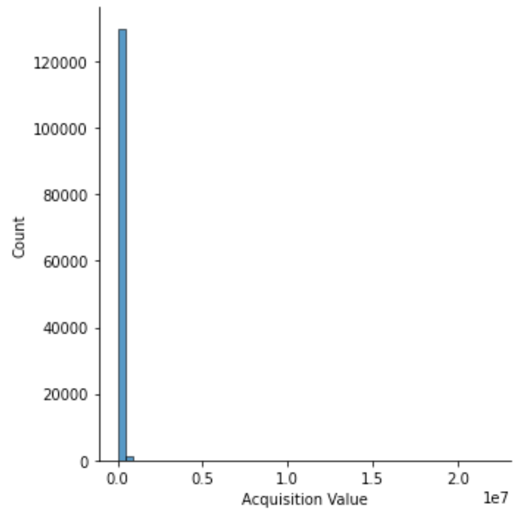


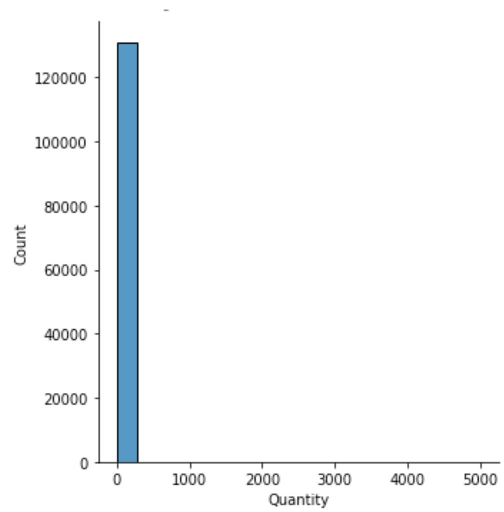
- DEMIL IC 1 makes up 85% of the transfers and most DEMIL Codes belong to it.
- DEMIL IC 6 is only used for DEMIL Code Q.
- DEMIL IC -1 is mostly used for DEMIL Code A.



- The states AK, DE, GU, NE, NM, MP, PR, RI, VT, VI, and WY had the lowest number of items transferred to them.
- CA, TX, SC, CH and IL are the states with the highest number of transfers.

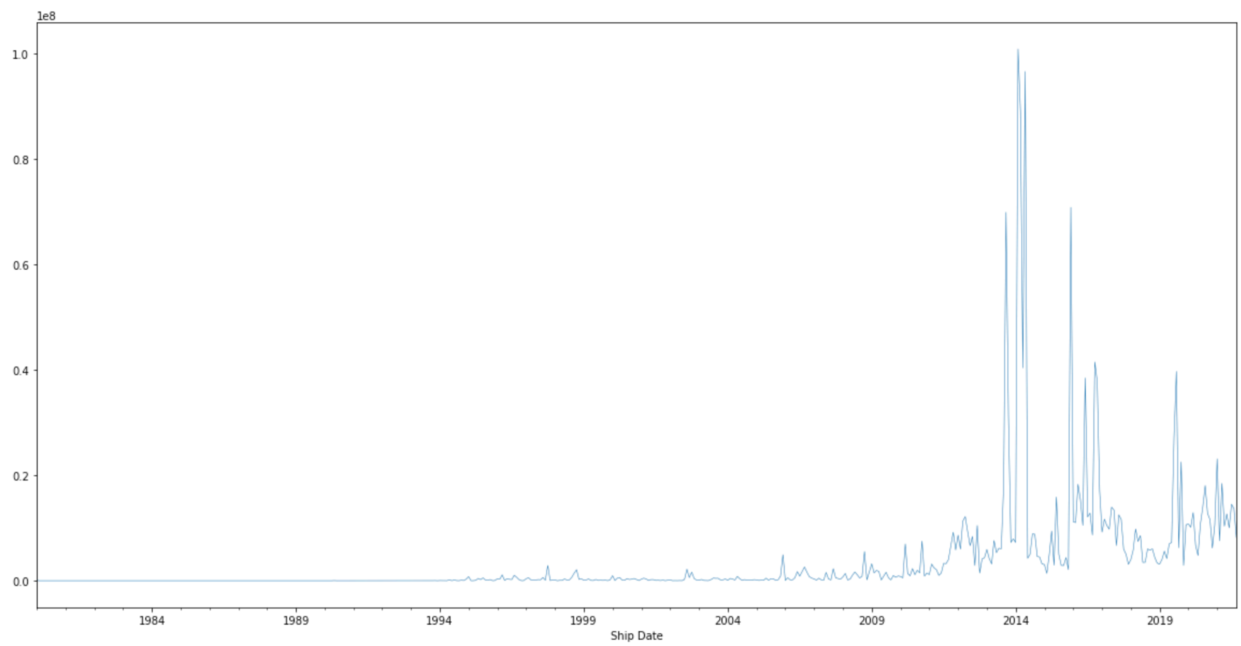
Acquisition Value, Quantity and Time Series Analyses



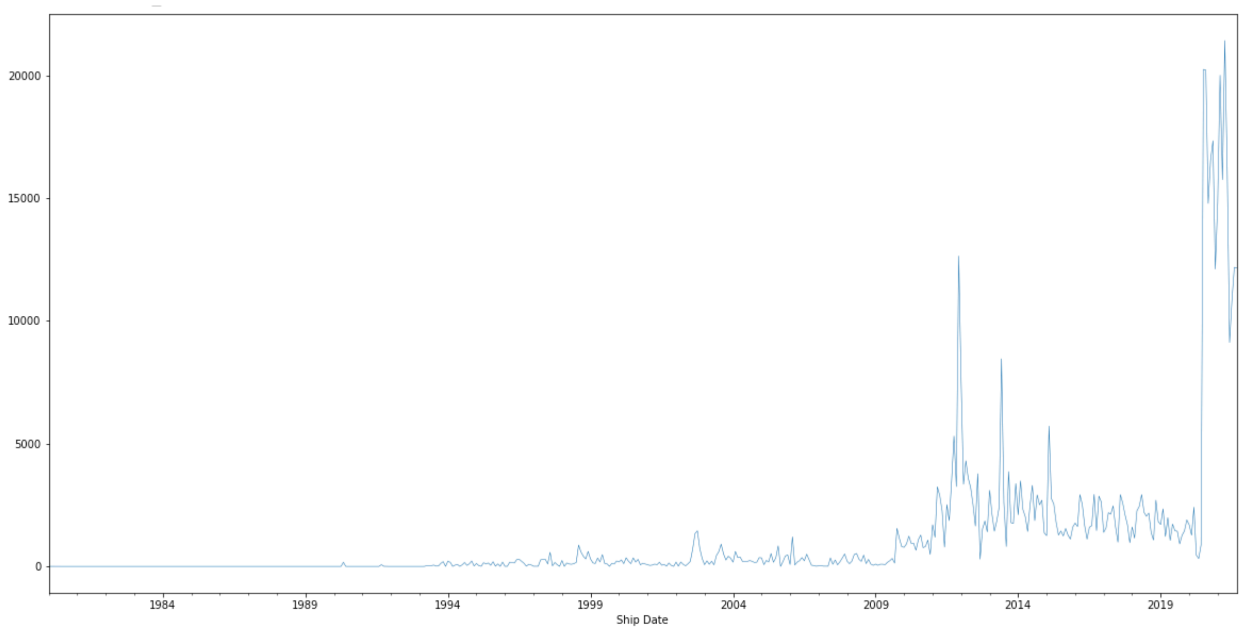


- Acquisition value and Quantity have extreme outliers as they are heavily right skewed.

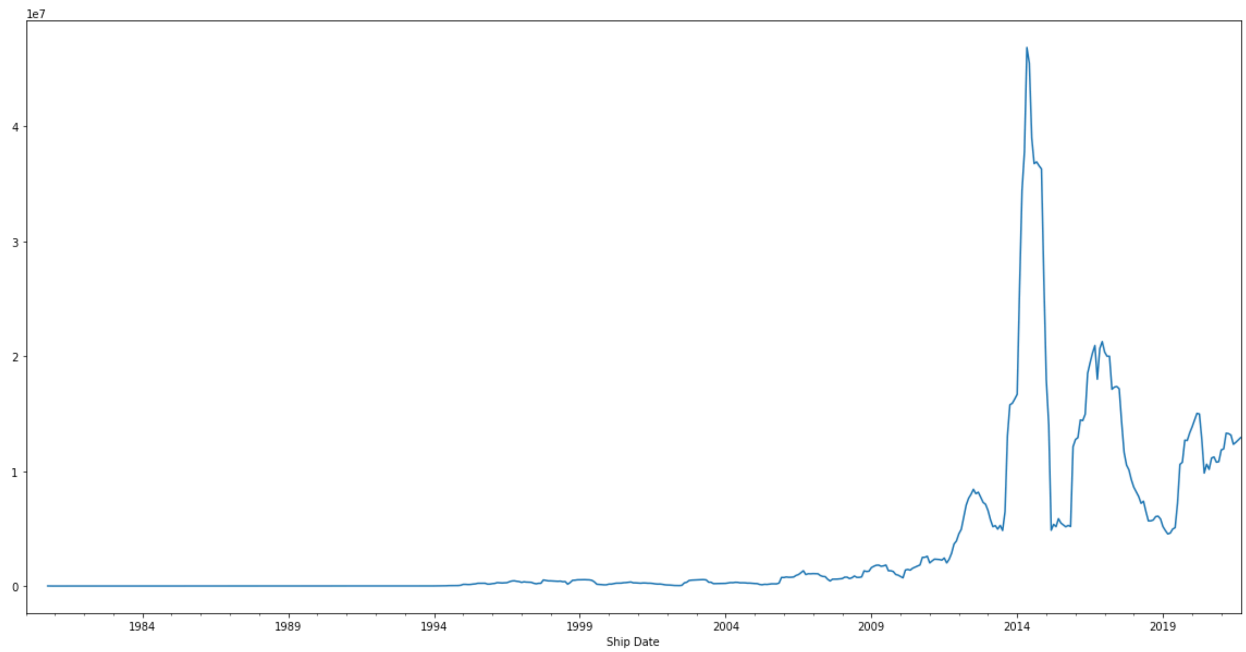
Time Series Analysis



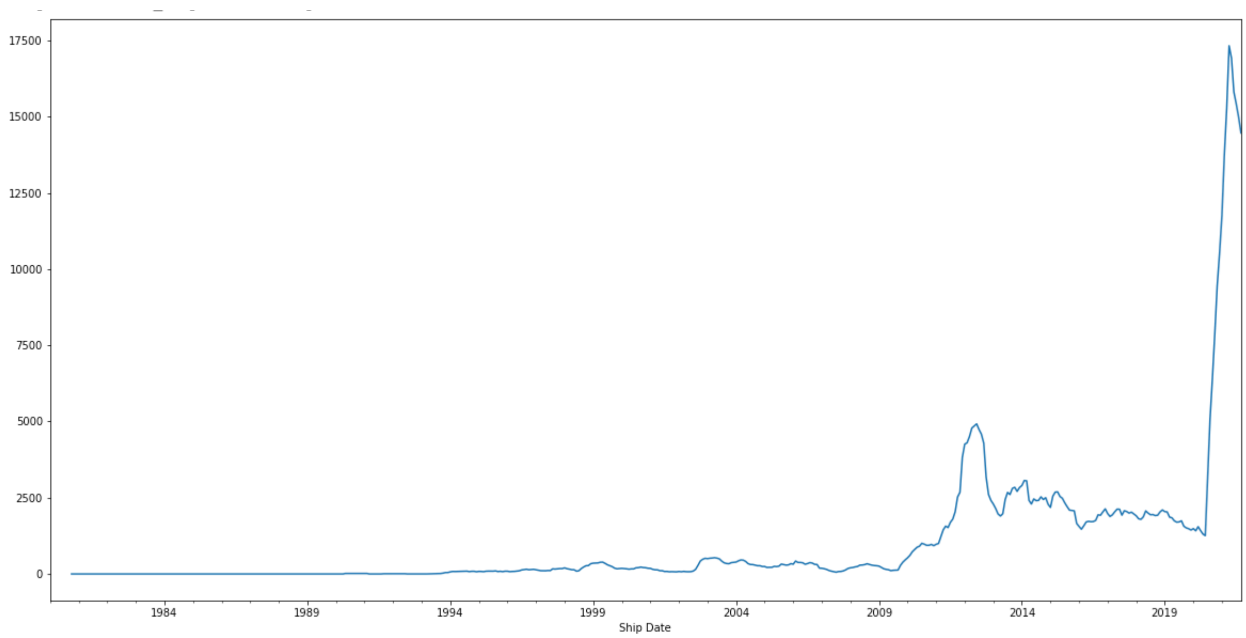
- There is no seasonality in the monthly acquisition values over the years.



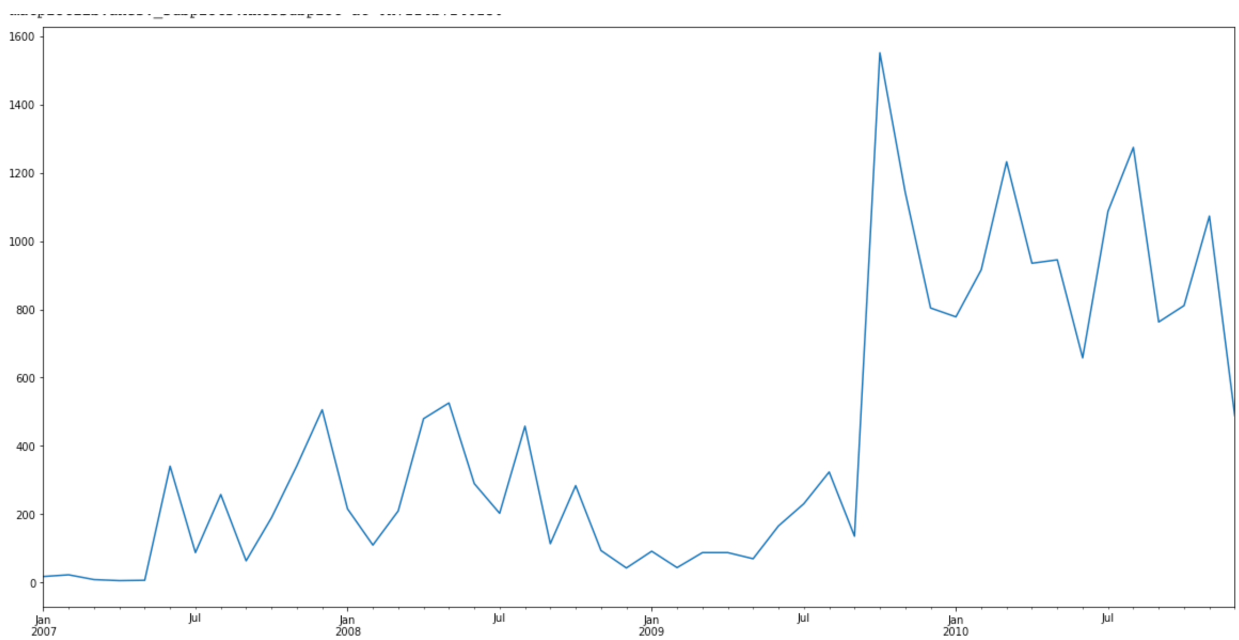
- There is no seasonality in the monthly quantity values over the years.



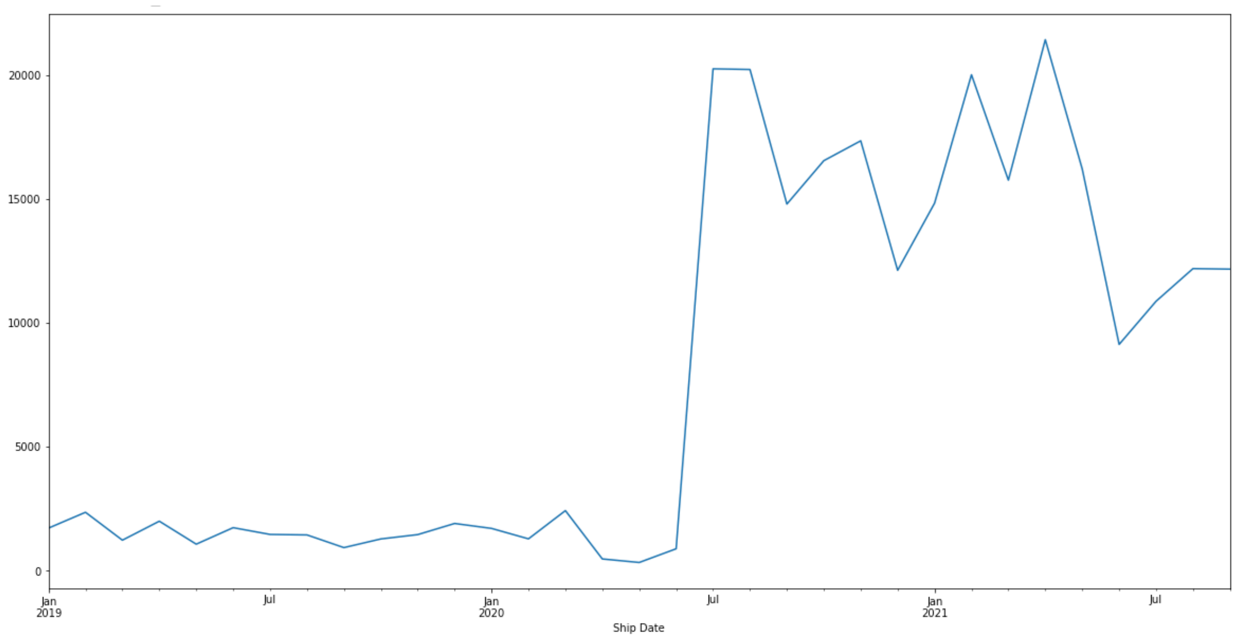
- Smoothing the acquisition value time series using simple moving average of order 10 better revealed the upward/positive trend of the data. Acquisition values transferred are increasing with time.



- Smoothing the quantity values time series using simple moving average of order 10 better revealed the upward/positive trend of the data. Item quantities transferred are increasing with time.



- Structural break in monthly quantity values happened in September 2009, mean quantities transferred increased after it.



- Structural break in monthly quantity values in June 2020, mean quantities transferred significantly increased after it.

State Related Analyses

The top and bottom agencies in each state in terms of number of transfers made to them, as well as the top items transferred to each state can be found in the notebook submitted with this report. Samples are shown below of each:

Top Agencies:

State	Agency Name	
AK	ANCHORAGE POLICE DEPARTMENT	113
	ALASKA DEPT OF PUBLIC SAFETY	73
	JUNEAU POLICE DEPARTMENT	15
	HAINES BOROUGH POLICE DEPT	5
	FAIRBANKS POLICE DEPT	1
AL	TUSCALOOSA POLICE DEPT	595
	DALE COUNTY SHERIFF OFFICE	234
	OXFORD POLICE DEPT	212
	TUSCUMBIA POLICE DEPT	179
	HOMEWOOD POLICE DEPT	141

Bottom Agencies:

State	Agency Name	
AK	FAIRBANKS POLICE DEPT	1
	NORTH POLE POLICE DEPT	1
	HAINES BOROUGH POLICE DEPT	5
	JUNEAU POLICE DEPARTMENT	15
	ALASKA DEPT OF PUBLIC SAFETY	73
AL	BULLOCK COUNTY SHERIFF DEPT	1
	DAUPHIN ISLAND POLICE DEPT	1
	ELMORE COUNTY SHERIFF OFFICE	1
	GARDENDALE POLICE DEPT	1
	HAYDEN POLICE DEPARTMENT	1

Top Items of each State:

State	Item Name	
AK	RIFLE,5.56 MILLIMETER	70
	RIFLE,7.62 MILLIMETER	13
	Mittens, cold weather	8
	LINER,WET WEATHER PONCHO	5
	TARPAULIN	5
	UNMANNED VEHICLE,GROUND	5
	GLOVES,COLD WEATHER	4
	TROUSERS,EXTREME COLD WEATHER	4
	PAPER,TOILET	3
	SLEEPING BAG	3
AL	RIFLE,5.56 MILLIMETER	1388
	PISTOL,CALIBER .45,AUTOMATIC	417
	RIFLE,7.62 MILLIMETER	332
	SIGHT,REFLEX	327
	ILLUMINATOR,INFRARED	230
	TRUCK,UTILITY	179
	NIGHT VISION GOGGLE	116
	IMAGE INTENSIFIER,NIGHT VISION	95
	MAGAZINE,CARTRIDGE	85
	WEAPON PARTS	74

Top States with Highest Quantities Ordered in Each Year (sample):

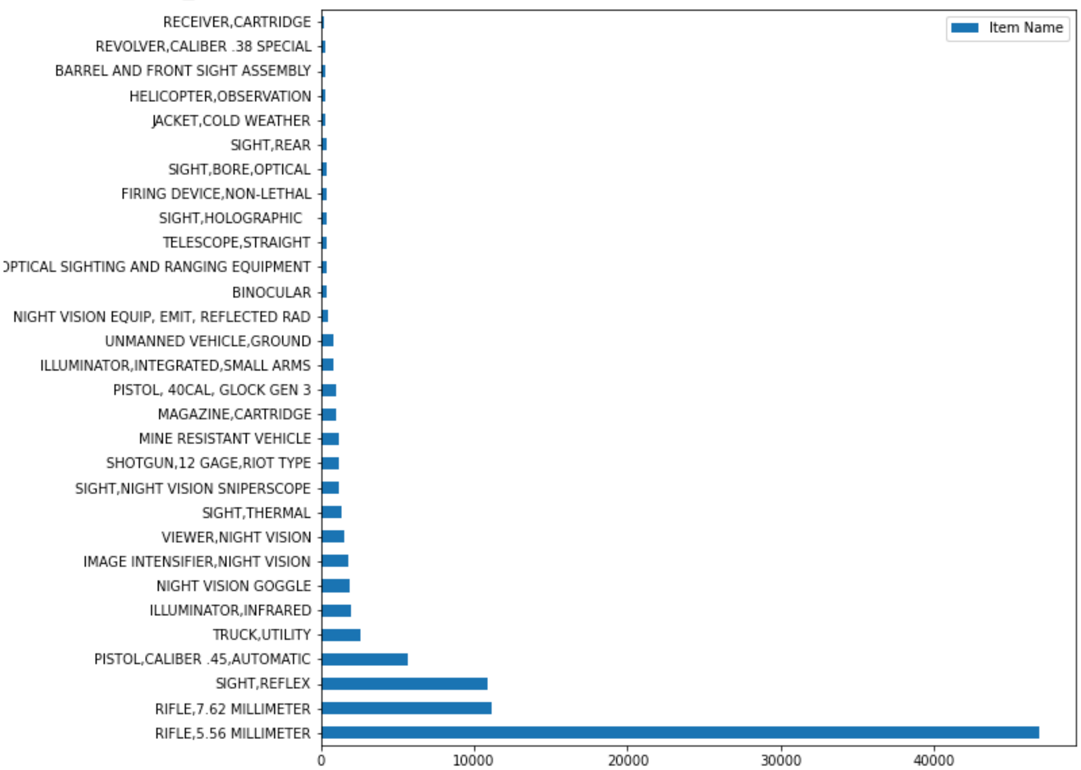
Ship Date	State	Quantity
1980	NY	1
1990	MT	167
1991	WY	69
	MO	4
1992	AZ	3
1993	KY	90
	SD	63
	CT	48
	ND	36
	NH	28
1994	OH	105
	CO	84
	RI	76
	IL	74
	LA	74

Top States with Highest Acquisition Value in Each Year (sample):

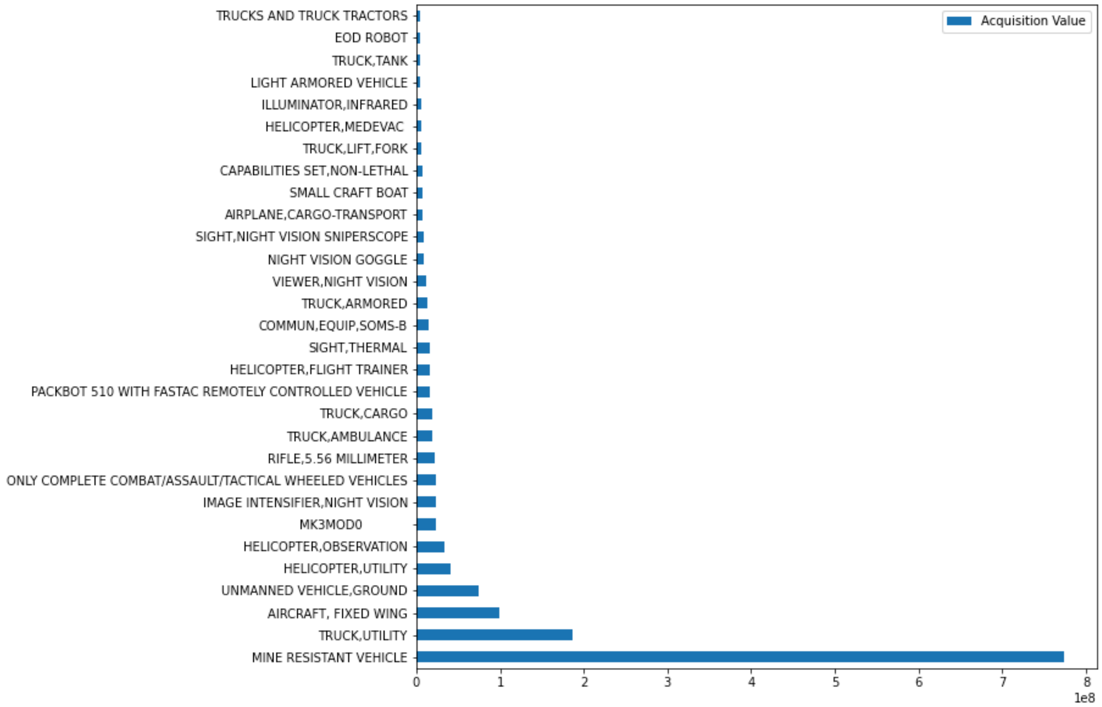
Ship Date	State	Acquisition Value
1980	NY	65070.00
1990	MT	23046.00
1991	WY	9522.00
	MO	552.00
1992	AZ	7261.80
1993	KY	12420.00
	SD	8694.00
	CT	6624.00
	WY	5255.00
	ND	4968.00

Item Related Analyses

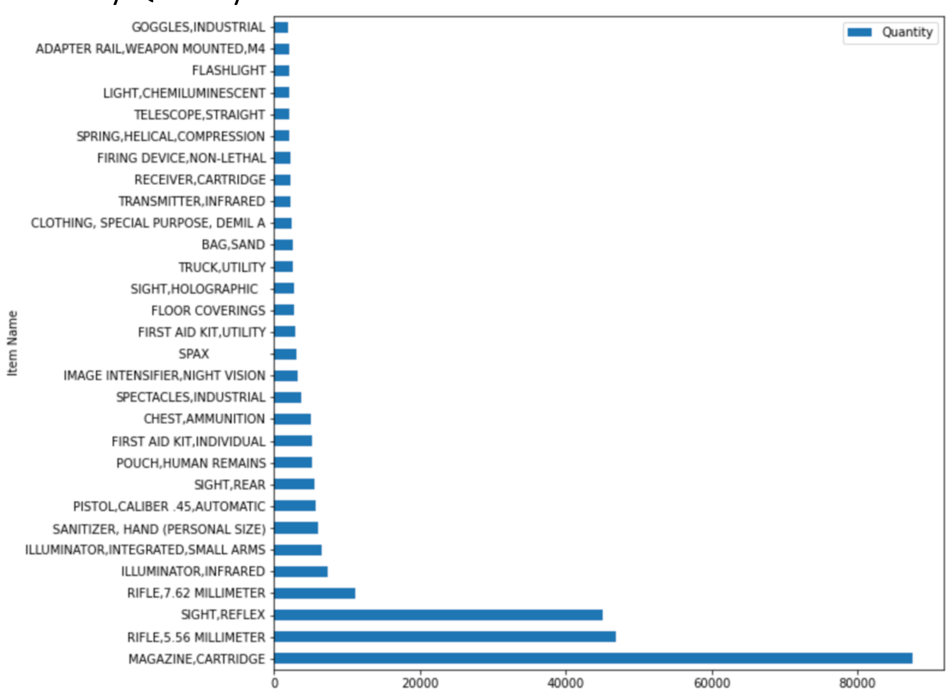
Top Overall Ordered/transferred Items:



Top Items Ranked by Acquisiton Values Transferred:



Top Items Ranked by Quantity Transferred:



DEMIL Code, DEMIL IC and other Attributes

Distribution of DEMIL IC Values Over DEMIL Code Values:

DEMIL IC DEMIL Code	-1.0	0.0	1.0	3.0	4.0	5.0	6.0	7.0
A	0.633502	0.023810	0.103892	0.000000	0.000000	0.0	0.0	0.121411
B	0.011806	0.023810	0.000000	0.007029	0.000000	0.0	0.0	0.000000
C	0.054188	0.023810	0.055414	0.000000	0.285714	0.0	0.0	0.082670
D	0.294753	0.857143	0.795303	0.000000	0.714286	0.0	0.0	0.374957
E	0.000303	0.000000	0.001183	0.000000	0.000000	0.0	0.0	0.002767
F	0.002624	0.000000	0.044138	0.000000	0.000000	0.0	0.0	0.418194
Q	0.002825	0.071429	0.000071	0.992971	0.000000	1.0	1.0	0.000000

- 63% of non-reviewed DEMIL Codes (DEMIL IC -1) are code 'A'.
- DEMIL IC 0 and 1 are mostly for DEMIL Code 'D'.
- DEMIL IC 3 is 99% for DEMIL Codes 'Q'. and DEMIL IC 6 is all for DEMIL Codes 'Q'. This is expected given the definition of 'Q'.
- All DEMIL IC 5 are for DEMIL Code 'Q'.
- 42% of DEMIL IC 7 are for DEMIL Code 'F'.

Distribution of DEMIL Code Values Over Different DEMIL IC Values:

DEMIL IC DEMIL Code	-1.0	0.0	1.0	3.0	4.0	5.0	6.0	7.0
A	0.342835	0.000055	0.637942	0.000000	0.000000	0.000000	0.000000	0.019168
B	0.750000	0.006410	0.000000	0.243590	0.000000	0.000000	0.000000	0.000000
C	0.076583	0.000143	0.888620	0.000000	0.000570	0.000000	0.000000	0.034084
D	0.031248	0.000385	0.956664	0.000000	0.000107	0.000000	0.000000	0.011596
E	0.020833	0.000000	0.923611	0.000000	0.000000	0.000000	0.000000	0.055556
F	0.004195	0.000000	0.800742	0.000000	0.000000	0.000000	0.000000	0.195063
Q	0.004949	0.000530	0.001414	0.948745	0.000000	0.000177	0.044185	0.000000

- 34% of code 'A' and 75% of code 'B' were not reviewed (DEMIL IC -1).
- The rest of the codes excluding 'Q' mostly had IC of 1. Most transfers possess high integrity.
- 95% of code 'Q' had IC of 3. This means that 95% of code 'Q' items require mutilation inside of the US too.

DEMIL Code Ordered by Total Acquisition Value Transferred:

DEMIL Code	Acquisition Value
E	3.426224e+04
B	2.351485e+06
F	5.618198e+07
A	8.865112e+07
Q	1.840856e+08
D	1.968024e+08
C	1.195331e+09

DEMIL IC Ordered by Total Acquisition Value Transferred:

DEMIL IC	Acquisition Value
5.0	3.347000e+03
4.0	2.551941e+05
0.0	3.550394e+05
6.0	2.528442e+07
7.0	6.423735e+07
3.0	1.589664e+08
-1.0	2.014019e+08
1.0	1.272934e+09

- Items with non-reviewed DEMIL Codes have the second highest acquisition value transferred.

DEMIL Code Ordered by Total Quantity Ordered:

DEMIL Code	Quantity
E	803
B	1040
F	7999
C	11940
Q	26871
A	218990
D	279792

- Code ‘A’ items, which are mostly non reviewed items, have the second highest quantities transferred/ordered (by local agencies).

DEMIL IC Ordered by Total Quantity Ordered:

DEMIL IC	Quantity
5.0	4
4.0	21
0.0	368
6.0	870
7.0	8320
3.0	26097
-1.0	64779
1.0	446976

- Items with non-reviewed DEMIL Codes have the second highest quantities transferred.

Data Cleaning and Feature Engineering

To clean the data, DEMIL IC was converted to a string attribute and all missing values were replaced with ‘Not Reviewed/No Integrity’. After that, records with erroneous zeros were removed. These are the records with acquisition value or quantity of 0. Realistically, if a transfer happened then some quantity must have been transferred. Moreover, every item has some acquisition value associated with it no matter how small.

New attributes were derived from the ship date attribute to make better use of it. These attributes were ship day, ship month and ship year. These attributes are expected to help the models in learning information about the data that is otherwise impossible to learn without using date information.

Agency name, NSN, UI, station type and ship date were removed. Agency name was removed as it is expected to give redundant information given that each agency is unique to a state and the state attribute will be used. Agency name also has very high cardinality (1402 unique values). For the same reasons, NSN was removed as it represents the equipment code, but the attribute Item Name is present and like NSN, it represents the item transferred. Also, NSN has very high cardinality (31543 unique values). UI was removed as 125920 records have UI as 'Each', meaning 96% of the dataset have the same value for UI; no useful information can be learned from it. Similarly, station type was removed as it has one unique value only. Lastly, ship date was removed as it cannot be used in the format it is in and features were already derived from it.

Outliers in numeric attributes were not removed as neural networks can handle outliers well and although outliers are present, their percentage in the dataset is very small.

The dataset was split to a training and testing sets, the training set was further split to training and validation sets. For all sets, features were transformed by scaling numerical features to be within a range of 0 to 1 and categorical features were transformed by one-hot encoding them. For all sets, the target was label encoded.

Description of Methods Used

Three types of models were developed to predict DEMIL Code. Firstly, a random forest model was built and used as a baseline and benchmark for assessing performance. Secondly, artificial neural network models (no hidden layers and single hidden layer networks) were built. Lastly, a deep neural network model was built.

The two ANN models used different representations of categorical data in making predictions. The first ANN used one-hot encoded features as its input. The second ANN used embeddings of the categorical features as input. In contrast to one-hot encoding, word embeddings are condensed representations of words/categories. Each word is represented by a vector where all elements of the vector are used to represent the word. The number of elements in the vectors can be significantly less than the number of categories present while still being able to represent all categories.

Feature embeddings are created in the following way/have the following attributes:

1. Number of unique values in the feature (length of the vocabulary of the feature) is used to create a lookup table of that length where each row in it has

    a. An index which is an integer (one index corresponds to one instance from the vocabulary).
    b. The corresponding embedding.
2. Each embedding is a vector that represents the integer index.
3. All embedding vectors have the same size.
4. The length of the vocabulary is considered as the input length of the embedding.
5. The size of each embedding vector is considered as the output length of the embedding.

Using one-hot encoding allows a model to memorize feature specific combinations, which helps in prediction but at the detriment of generalization. In contrast, the dense and low dimensional representations of features using embeddings enable a model to generalize better. The model would generalize better as it can match new data points based on their closeness in the embedding space to previously seen data.

The DNN model used both approaches. It used both one-hot encoding and embedding representations of the categorical features to make predictions. That was done by making the DNN model have two parts, one part which has three hidden layers and uses the one-hot encoding representation of the features as its input and another part which also had three hidden layers but takes the embedding representations of the features as its input. The results of the two parts are merged before being passed to the output layer.

As this is an imbalanced classification problem, the DNN model was also trained using custom class weights to combat the imbalanced class distribution and its effects on classification. The class weights chosen were based on the percentages of the classes in the dataset. The weight of each class was set to be 1 / percentage of the class in dataset. This way, the smaller the percentage, the higher the weight will be. This gives underrepresented classes way higher weights.

Results

The performances of the models were as follows:

Random Forest (Benchmark Model)

	precision	recall	f1-score	support
0	0.97	1.00	0.99	3642
1	0.75	0.36	0.49	33
2	0.98	0.89	0.93	1386
3	0.99	0.99	0.99	18529
4	1.00	0.36	0.53	22
5	1.00	0.97	0.98	1229
6	0.99	0.98	0.99	1131
accuracy			0.99	25972
macro avg	0.95	0.79	0.84	25972
weighted avg	0.99	0.99	0.99	25972

The random forest performed very well in terms of accuracy and precision across the classes. However, the recall of the underrepresented classes (1 and 4) is considerably lower than that of all other classes. To be able to justify using neural network models, the NN models need to perform better than this random forest model as a random forest is less computationally demanding and more easily interpretable.

ANN Using One-Hot Encoding

	precision	recall	f1-score	support
0	0.96	0.97	0.96	3642
1	0.00	0.00	0.00	33
2	0.99	0.62	0.76	1386
3	0.94	1.00	0.97	18529
4	0.00	0.00	0.00	22
5	0.99	0.71	0.83	1229
6	1.00	0.74	0.85	1131
accuracy			0.95	25972
macro avg	0.70	0.58	0.62	25972
weighted avg	0.95	0.95	0.94	25972

The ANN using one-hot encoding performed well in terms of accuracy and precision across the classes besides classes 1 and 4 which it could not predict at all. This model performed worse than the benchmark model.

ANN Using Embeddings

	precision	recall	f1-score	support
0	0.95	0.98	0.97	3642
1	1.00	0.36	0.53	33
2	0.81	0.82	0.82	1386
3	0.98	0.98	0.98	18529
4	1.00	0.36	0.53	22
5	0.88	0.86	0.87	1229
6	0.97	0.89	0.93	1131
accuracy			0.96	25972
macro avg	0.94	0.75	0.80	25972
weighted avg	0.96	0.96	0.96	25972

The ANN using embeddings performed well in terms of accuracy and precision across the classes. This ANN has also produced way higher recall scores across the classes besides class 1 and 4 for which it had the same recall scores as the benchmark. This model has outperformed the benchmark model in producing higher recall scores.

DNN

	precision	recall	f1-score	support
0	0.96	0.99	0.98	3642
1	1.00	0.61	0.75	33
2	0.89	0.85	0.87	1386
3	0.99	0.99	0.99	18529
4	0.92	0.50	0.65	22
5	0.89	0.92	0.91	1229
6	0.99	0.99	0.99	1131
accuracy			0.98	25972
macro avg	0.95	0.83	0.88	25972
weighted avg	0.98	0.98	0.98	25972

The DNN model which uses both one-hot encoding and embeddings has outperformed all models. It has produced high precision, recall and accuracy scores. Its precision across the classes is only slightly worse than that of the benchmark and is higher than that of other ANNs. Its recall scores are the best recorded so far.

DNN Using Custom Class Weights

	precision	recall	f1-score	support
0	0.97	0.97	0.97	3642
1	0.16	0.88	0.27	33
2	0.74	0.91	0.82	1386
3	1.00	0.96	0.98	18529
4	0.28	0.64	0.39	22
5	0.84	0.98	0.91	1229
6	0.99	0.98	0.99	1131
accuracy			0.96	25972
macro avg	0.71	0.90	0.76	25972
weighted avg	0.97	0.96	0.96	25972

Using the custom class weights has significantly improved the model's recall scores for the underrepresented classes 1 and 4 to be like those of the other classes. That came at the cost of worsening precision.

Predicting the DEMIL Code is an output sensitive problem where it is more important to have less false negatives than it is to have less false positives. For example, since DEMIL Code 'B' means mutilation to the point of scrap is required worldwide, falsely predicting that a DEMIL Code is 'B' (false positive) is less bad than falsely predicting an actual DEMIL Code 'B' as another code (false negative) that could potentially represent a lower demilitarization degree. Because of that, having higher recall scores is more important and thus the DNN that uses custom class weights is the best performing model.

Summary and Insights

Best Predictor Model

The best model for prediction is a deep neural network that uses both one-hot encoding and embedding representations of the input to make predictions as shown and discussed in the results section previously.

Patterns Found in the Data

- Quantities of items ordered by agencies/transferred to agencies and the amount of acquisition values transferred are in an upward trend. Demand is only increasing.
- Three out of the five topmost transferred items are weapons (rifles and pistols), indicating increased militarization of local police enforcements.
- 63% of non-reviewed DEMIL Codes (DEMIL IC -1) are code 'A'.
- 34% of code 'A' and 75% of code 'B' were not reviewed (DEMIL IC -1).
- 42% of DEMIL IC 7 are for DEMIL Code 'F'.

Recommendations Based on Findings

- It is recommended to increase scrutiny on the integrity of the demilitarized code as items that are subject to the least demilitarization (code A) are the ones with lowest integrity checks (not reviewed, no DEMIL IC provided).
- It is recommended to investigate why most codes that ICP has not responded to collaboration request in changing them (over 90 days old) or failed to update them in the ICPs legacy system (this is the description of DEMIL IC 7) are codes of class F.
- Since one of the KPIs of the LDA is to increase service readiness and supply availability of its agencies (as found on the LDA's website), it is recommended to survey the states as well as the agencies across the states that have the least number of transfers/orders associated with them about the reasons behind them only ordering few equipment/items and how can the LDA services be improved to enable them to order more and better meet their needs.

Special Cases to Consider

- Structural changes have happened in the gross quantities ordered/transferred in both September 2009 and June 2020. These structural changes should be further analyzed to uncover reasons behind these structural changes to be more ready for the next structural change and the increase in demand (quantities ordered) that is anticipated to come after it.

Bibliography

Brownlee, J. (2019). 3 Ways to Encode Categorical Variables for Deep Learning. [online] Machine Learning Mastery. Available at: https://machinelearningmastery.com/how-to-prepare-categorical-data-for-deep-learning-in-python/.

fchollet (2019). Keras documentation: Imbalanced classification: credit card fraud detection. [online] keras.io. Available at: https://keras.io/examples/structured_data/imbalanced_classification/

Heaton, J. (2019). What are Embedding Layers in Keras (11.3). [online] www.youtube.com. Available at: https://www.youtube.com/watch?v=OuNH5kT-aD0

Heng, C. (2016). Wide & Deep Learning: Better Together with TensorFlow. [online] Google AI Blog. Available at: https://ai.googleblog.com/2016/06/wide-deep-learning-better-together-with.html

Salama, K. (2020). Keras documentation: Structured data learning with Wide, Deep, and Cross networks. [online] keras.io. Available at: https://keras.io/examples/structured_data/wide_deep_cross_networks/#experiment-2-wide-amp-deep-model

TensorFlow (n.d.). Classify structured data using Keras preprocessing layers. [online] TensorFlow. Available at: https://www.tensorflow.org/tutorials/structured_data/preprocessing_layers