Predictive Word Patterns in Medical Device Adverse Events

A Thesis

By Harry Goldman

Northwestern University

School of Professional Studies

Masters of Science in Predictive Analytics Program

May 2015

Table of Contents

Introduction

This work attempts to solve a classification problem in medical device adverse event reporting by mining the text provided in the report narrative describing the event and then classifying the reports based on a predictive model. Adverse event reports (reports) are typically classified as one of: "Death", "Injury", "Malfunction" or "Other". Not all reports are classified correctly with the most frequently misclassified occurring in the "Other" category. The narrative text consists of short descriptions, medical terminology and abbreviations. Using the event narrative to predict the correct event type can provide a better classification system for adverse event reports by automating the classification process.

Event narratives are similar to maintenance or machine logs: free-form text with a great deal of variation and no real standards. In the information extracted "all of the expressive freedoms possible in natural language with the additional freedoms of not having to adhere to the usual constraints of spelling, grammar, or vocabulary" (Mark Devaney, 2005) is an apt definition of the corpus of words and phrases associated with adverse events and the original classification of the event.

The resulting corpus exhibits many characteristics which make standard approaches to natural language processing and text classification difficult:

- The phrases utilized in the event narratives are not used consistently
- The phrases used do not always correspond to a medical device or medical procedure
- The event narrative is free-form text often with short or abbreviated and fragmented sentences with poor sentence structure
- Medical terms and abbreviations can make understanding the narrative difficult
- Abbreviations are often used which may not always map clearly to the intended word or phrase
- Medical jargon and other words which may be ambiguous when taken out of context

Data – Overview

Data used in this project was downloaded from the FDA database of medical device adverse events (MDRs) for the year 2011. The data is publicly available on the FDA web site. The FDA makes the following comments regarding medical device adverse events:

"Each year, the FDA receives several hundred thousand medical device reports (MDRs) of suspected device-associated deaths, serious injuries and malfunctions. The FDA uses MDRs to monitor device performance, detect potential device-related safety issues, and contribute to benefit-risk assessments of these products. The MAUDE database houses MDRs submitted to the FDA by mandatory reporters 1 (manufacturers, importers and device user facilities) and voluntary reporters such as health care professionals, patients and consumers.

Although MDRs are a valuable source of information, this passive surveillance system has limitations, including the potential submission of incomplete, inaccurate, untimely, unverified, or biased data. In addition, the incidence or prevalence of an event cannot be determined from this reporting system alone due to potential under-reporting of events and lack of information about frequency of device use. Because of this, MDRs comprise only one of the FDA's several important postmarket surveillance data sources." (U.S. Food and Drug Administration, 2014)

Figure 1 (below) shows the frequency of medical device adverse event reports for the year 2011 with breakdowns by month and type. While "Death" and "Injury" reports are of the greatest concern to device manufacturers a review of all incoming reports is required. Reports categorized as "Other" must be properly categorized and all reports must be reviewed to identify miscategorized reports. Reports classified as "Other" (Type "O" or "*") will have an unknown outcome and as such will require review and analysis on receipt by the device manufacturer. This is a timely and costly manual process.

It should be noted that any year could have been chosen for this work, the year 2011 was selected randomly from the years 2008 through 2013. The sole criteria was the data for the year must be complete, which was true for all candidate years at the time the data was downloaded from the FDA website.

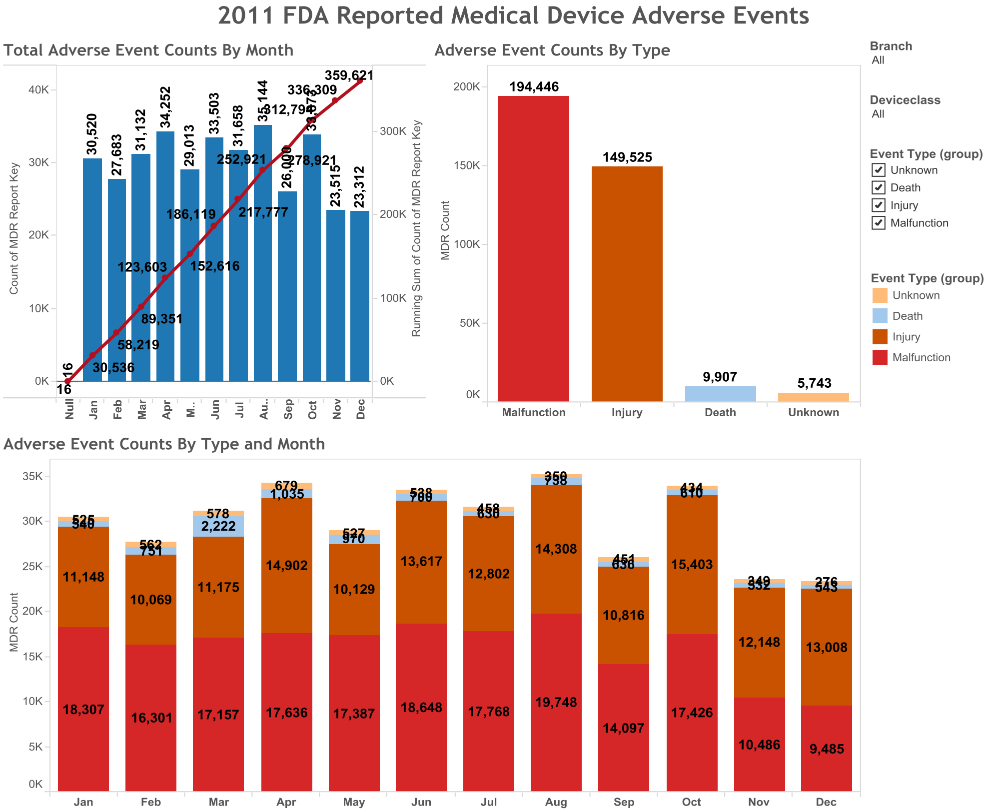


Figure 1- 2011 Medical Device Adverse Events

Related Work

Text analytics is a rich topic with many published papers but the specific area of medical device adverse events has seen very little activity. Parallels exist between MDR reports and various types of trouble tickets or maintenance reports and this is where the majority of similar scholarly research has been. Of particular interest for adverse event analysis is the area of sublanguages. Adverse event reporting uses its own sublanguage allowing for classification of terms and synonyms, similar to work done in areas such as maintenance log reviews.

Liddy (Elizabeth D. Liddy C. L., 1993) discusses the sublanguage approach when she states "First, the sublanguage approach is similar to discourse analysis in that it takes advantage of the fact that particular language usage patterns develop within the written texts or oral communications of a community that uses this sublanguage either to accomplish some common task or to discuss some common activity. A sublanguage has a restricted vocabulary which is distinctive in the set of words that comprise it and may have some rather unusual, perhaps ungrammatical syntactic characteristics.

"…we suggest that the grammar of a sublanguage reflects the information structure of discourse in the subfield, whereas the semantic classes of words used and the semantic relations between these classes reflect the knowledge structure of a subfield." (Elizabeth D. Liddy C. L., 1993)

MDR reports have their own language - a blend of medical jargon with abbreviations and short terse sentence structure. Examples of the narrative text found:

MDR Report	Date Received by FDA	Report Text
1952681	1/4/2011	ACCOUNT ALLEGED THAT THE POWER SUPPLY BOARD IS CORRODED DUE TO FLUID INGRESS. ACCOUNT STATED THAT THERE WERE NO INJURIES AND A PT WAS NOT IN THE BED.
1949440	1/5/2011	THE CUSTOMER STATES THAT THE INTEGRATED CHIP TECHNOLOGY (ICT) CUP AND MODULE ARE FAILING TO FLUSH ON THE ARCHITECT C8000 ANALYZER. THE BULK SOLUTIONS ARE ALSO FAILING TO FLUSH. THE ICT REFERENCE SOLUTIONS ARE FAILING TO ASPIRATE. CONTROLS AND CALIBRATIONS ARE FAILING. A SERVICE CALL WAS INITIATED. THE CUSTOMER CALLED BACK AND STATES THAT ISSUE CONTINUES. THE CUSTOMER IS REQUESTING SERVICE TO RETURN. THE FIELD SERVICE ENGINEER (FSE) NOTICED THAT THE ICT PROBE WAS MISALIGNED AND HITTING THE SIDE OF THE REFERENCE SOLUTION CUP. THE FSE REALIGNED THE PROBE. HOWEVER, THE FOLLOWING DAY, THE CUSTOMER CALLED AND STATED THAT THE ISSUE CONTINUES WITH THE SODIUM ASSAY. ONE PATIENT SAMPLE GENERATED AN INITIAL RESULT OF 111 MMOL/L THAT RETESTED AT 140 MMOL/L. NO SUSPECT RESULTS WERE REPORTED FROM THE LAB. THE CUSTOMER REPLACED THE ICT MODULE AND PROBE WITH NO IMPROVEMENT. THE CUSTOMER WAS SENT AND INSTALLED A NEW ICT SYRINGE VALVE TUBING AND WILL MONITOR RESULTS. THERE IS NO IMPACT TO PATIENT MANAGEMENT REPORTED.
2061636	4/18/2011	IT WAS REPORTED THAT THE PROCEDURE WAS TO TREAT RE-STENOSIS OF UNKNOWN STENTS THAT WERE IMPLATED 8 YEARS PRIOR. THE PROMUS STENT WAS SUCCESSFULLY IMPLANTED ON (B)(6) 2011 FOR TREATMENT. ON (B)(6) 2011, THE PATIENT EXPERIENCED CARDIAC ARREST, AND EXPIRED. THE PHYSICIAN NOTED THAT THE DEATH MAY HAVE BEEN A RESULT OF ELECTRICAL INTERFERENCE. NO ADDITIONAL INFORMATION WAS PROVIDED.

The sublanguage uses a subset of the normal lexicon of language leaving large parts of the natural language ignored. This includes the various tenses and senses of words which may otherwise need to be identified and handled in the classification process. Further a sublanguage will typically have a smaller number of verbs and nouns than the full language possesses which may help in the classification of terms by allowing a better understanding of the text of interest as words which may have more than one meaning will have a much clearer purpose in a sublanguage. “At the syntactic level, a sublanguage is characterized by a limited range of verbs and makes extensive use of nominal compounds which reflect the specialized nature of the subfield. In addition, a sublanguage is frequently telegraphic in nature, dropping articles, auxiliary verbs, and even conjunctions.” (Elizabeth D. Liddy C. L., 1993)

Similar work was done by Devaney: “The objective is to extract and categorize machine components and subsystems and their associated failures using a novel approach that combines text analysis, unsupervised text clustering, and domain models. Through industrial partnerships, this project will demonstrate effectiveness of the proposed approach with actual industry data.” (Mark Devaney, 2005)

Devaney’s work focused on “discovering knowledge required for critical business performance improvements by mining the equipment maintenance logs collected by the owners of complex machinery.” (Mark Devaney, 2005) The raw data used in Devaney’s work was similar in structure to MDR reports; his data had a series of structured fields defining the incident and an unstructured text field containing the narrative.

Devaney identifies several challenges which made classification more difficult:

- “Vocabulary used in descriptions is inconsistent. There is typically no standard set of terms used for the names of mechanical parts or the activities performed on them.
- Vocabulary may not correspond directly to systems or components of interest.
- Input is not well-formed. Because of character length limitations and probable treatment of the log entries as secondary task, the text descriptions are limited to short phrases or sentence fragments. Additionally, attention is not typically paid to spelling and other

language rules, and the data therefore exhibits very high degrees of grammatical and spelling errors.

- Jargon and extremely terse abbreviations are common. Due to the limited input space and time pressures of maintenance engineers, log entries are usually laced with creative abbreviations and contain large amounts of jargon.
- Large amounts of data are not available. … Thus, a new type of machine put into service will have on the order of 50-100 maintenance actions performed per year. While this is too large for manual analysis, it is too sparse for purely data-intensive approaches." (Mark Devaney, 2005)

Devaney's methodology utilizes both a text analyzer and a clustering engine. Both these functions are incorporated into the SPSS software utilized in this work. Delaney notes: "While text analytics represent an important component of this project, our primary focus is on the definition and construction of a powerful representational framework as described above and the development of a clustering algorithm to produce natural categories from the raw log data" (Mark Devaney, 2005) This is similar to the work done here where the analytic portion of the process provides a predictive element to incoming MDR reports.

The generalized task of content analysis is summarized by Krippendorff as: "Content analysis is a research technique for making replicable and valid inferences from data to their context." (Krippendorff, 1989) Krippendorff breaks the task of content analysis into six steps: Design, Utilizing, Sampling, Coding, Drawing inferences and Validation. (Krippendorff, 1989)

In this work the majority of these steps are completed with the IBM SPSS Modeler® version 16.0 (referred to here as "SPSS") software. The SPSS environment provides the user a GUI based environment and the software handles most of the core tasks of analyzing the data, extracting concepts and assists in the establishing of rules for the processing of the documents. The creation of a relevant corpus of words and concepts by creating synonyms and syntax rules around the concepts extracted by SPSS along with the creation and correction of categories was a necessary and significant effort.

Syntactical rules, synonyms and text rules were utilized to identify core terms, common elements and many of the medical terms found in the corpus. Many of these were originally categorized incorrectly by SPSS. For example:

- Blood glucose issues - categorized under carbohydrates in the "Drug" category
- Proximal Seal (a device used in cardiac surgery) - categorized as a fastener in the "Restraint" category
- Treatment Cycle - categorized under "Cycling" as a Sport

The SPSS environment provides the user the ability to determine features of the documents both visually and through rules. Yang and Pederson (Yiming Yang, 1997) discuss methods for feature selection and provide formulas for feature selection. The SPSS environment handles much of this work for the user. Working with feature selection the user must choose carefully the terms to include or exclude: "…. The basic assumption is that rare terms are either non-informative for category prediction, or not influential in global performance. In either case removal of rare terms reduces the dimensionality of the feature space. …. Also, DF is typically not used for aggressive term removal because of a widely received assumption in information retrieval. That is, low-DF terms are assumed to be relatively informative and therefore should not be removed aggressively. Information Gain (IG)" (Yiming Yang, 1997) In this work the incidence of low Documents Frequency (DF) providing Information Gain (IG) is low but relevant. Yang describes Document Frequency (DF) as "the number of documents in which a term occurs" and Information Gain (IG) as "the number of bits of information obtained for category prediction by knowing the presence or absence of a term in a document.) (Yiming Yang, 1997)

Yang and Pederson further note: "…IG measures the number of bits if information obtained by knowing the presence or absence of a term in a document. The strong DF-IF [DF-IG][1] correlations means that common terms are often informative, and vice-versa (this statement of course does not extend to stop words.)" (Yiming Yang, 1997)

---

[1] There appears to be a typo in the text by Yang and Pedersen, the term "DF-IF" in the text appears to be a typo and the correct term should be "DF-IG". Yang is referring to a strong correlation between Document Frequency (how often a term appears in the text) and Information Gain (information gained by knowing the term is present in the document.)

The SPSS software environment also handles most of the tasks outlined by Liddy when describing four stages of sublanguage analysis: Preprocessor, Lexical analysis, Adjacency analysis and Ambiguity filtering. In this work these stages were completed within the SPSS software allowing a more robust processing of the textual data. Some of these capabilities are not explored for this work as the objective of the work is to build a predictive model as opposed to adjacency analysis or determining the level of ambiguity found in each document.
In the work performed by Symonenko: “Analysis of a subset of tickets, guided by sublanguage theory, identified linguistic patterns, which were translated into rule-based algorithms for automatic identification of tickets’ discourse structure. The subsequent data mining experiments showed promising results, suggesting that sublanguage is an effective framework for the task of discovering the historical and predictive value of trouble ticket data.” (Svetlana Symonenko)

Like Symonenko, the task here is to utilize the text provided in the MDR data to not only determine the classification of the adverse event, but also to build a model to predict the classification types of incoming reports. Symonenko also touches on the sublanguage theme: “Trouble tickets exhibit a special discourse structure, combining system generated, structured data and free-text sections; a special lexicon, full of acronyms, abbreviations and symbols; and consistent “bending” of grammar rules in favor of speed writing” (Svetlana Symonenko)

In work published by Liddy (Elizabeth D. Liddy S. S.) on trouble tickets it was noted: “A feasibility study was conducted to determine whether the sublanguage methodology of NLP [Natural Language Processing] could analyze and represent the vital information contained in trouble tickets’ ungrammatical text and to explore various knowledge mining approaches to render the data contained in these documents accessible for analysis and prediction. Experiments showed that the linguistic characteristics of trouble tickets fit the sublanguage theoretical framework, thus enabling NLP systems to tap into the unrealized value of trouble ticket data.” (Elizabeth D. Liddy S. S.)

In the work on trouble ticket data by Liddy the domain specific linguistic attributes of the texts such as abbreviations, acronyms, and other semantic components were in line with other work in the sublanguage domain.

The SPSS software provides an environment for analyzing and categorizing concepts and phrases. In the SPSS environment challenges such as misspellings, changes in grammatical forms, abbreviations and short terse sentence structure are quickly and easily reduced to concepts and categories. Further the categorization process allows common categories to be combined and revised to correct missed meanings and types.

In the book *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications* Miner notes: "The goal of a text mining project to achieve a better predictive model or to identify more useful customer segments is to develop a deployable model that will turn text into numerical indices that can be used for automatic scoring (predictions). Thus, a common approach is to numericize the text using the methods just described. These methods accomplish various tasks, including the following:

1. Building a data matrix based on word or phrase counts or transformed counts
2. Computing various numeric indices based on those counts (relative frequencies, or document indices that capture, for example, grammatical complexity, or relevance determined a priori with respect to some target dimensions or categories
3. Merging of those indices (or raw counts) with other numeric indicators and information available for use in predictive modeling

The numericizing activities should be limited to only those features or dimensions of meaning that result from feature selection methods (see Chapter 9). The dimensions of meaning can be related to combinations of features, and both can be transformed (with both linear or nonlinear functions) to match their distributions. Often, this transformation operation improves the lift of the model (see Chapter 8)." (Gary Miner, 2012)

Data Processing

The steps used to build the candidate data file a listed here, and described below:

Step	Description	Records Affected	Records Remaining
1	Initial data file created		359,729
2	Drop records with no valid product code	108	359,621
3	Drop records for as Class 1 devices	13,237	346,384
4	Drop records with blank or NULL device class	2,977	343,407
5	Drop records with no reporting month	15	343,392
6	Select records for Orthopedic and Cardiology branches		152,969
7	Drop records with blank or NULL narratives	5,242	147,727
8	Select records for months 3,7,10 only		40,858

In the processing of the data, several issues were encountered: The data files in their original form are semi-normalized with a one-to-many relationship between the master record and both the device data and the narrative data. The combination of several files were necessary to produce a single source data set for the SPSS software. (The SPSS software could have potentially merged the various data files but the decision was made to handle this outside of SPSS.) The narrative text is contained in a character field with a maximum width of 255 characters. If the narrative text is longer than 255 characters the text is split into multiple rows. To properly capture the complete narrative text as a single document these records were combined into a single narrative record. The data files were processed in MS retaining only relevant fields from each joined file. The end result was a single file with 359,729 records. From this source all analysis, extractions and queries were performed.

From the initial data set 108 records were dropped because they did not have a valid product code, leaving a total of 359,621 records as candidates for analysis. Tables 1 and 2 (below) show the breakdown of the data by medical device class and by the month the FDA received the adverse event report.

Device Class	Count
1	13,237
2	194,436
3	148,971
(blank)	2,977
Grand Total	359,621

Table 1 - MDR Counts by Device Class

Month	Count
1	29,173
2	26,442
3	29,424
4	32,626
5	27,385
6	32,008
7	30,344
8	33,710
9	24,661
10	32,571
11	22,545
12	22,503
(blank)	15
Grand Total	343,407

Table 2 - Class 2 and 3 Counts by Month

From this pool of MDR reports any reports involving devices with a device class of 1 (least dangerous) were excluded as these devices rarely cause serious injury and even more rarely require investigation by the FDA. MDR reports were excluded based on the following criteria:

Criteria	Count
Device class 1	13,237
Device class blank	2,977
Report month blank	15
Total Excluded	16,229

After excluding these records the remaining data set of 343,392 MDR reports were considered the candidate data set for this study.

	Device Class			
Event Type	2	3	Grand Total	% of Total
M	137,439	44,691	182,130	53.04%
IN	51,815	94,738	146,553	42.68%
D	1,569	8,180	9,749	2.84%
O	1,974	1,015	2,989	0.87%
*	1,596	340	1,936	0.56%
(blank)	29	6	35	0.01%
Grand Total	194,422	148,970	343,392	

Table 3- MDR Reports by Event Type and Class

Table 3 (above) details the counts for MDR reports by event type and device class with just over half the events classified as "Malfunction" and 2.84% (9,749) classified as "Death" events.

For the medical device manufacturer the two biggest challenges with this data are:

1) Validation of the reported event type as correct
2) Properly classifying the "Other" and "*" events

Medical devices are assigned a product code which is used to identify broad classes of devices. Product codes are grouped by the branch of the FDA which provides oversight for those devices. Table 4 (below) provides a breakdown of the counts of reports by FDA branch and device class. A list of the top fifty product codes by report count with a breakdown of the event type, the FDA branch and the description of the device product code can be found in Appendix 1.

The top 50 product codes in Appendix 1 represents 271,934 MDR reports or 79.19% of all reports available for this study. Appendix 2 provides further analysis of the report data.

	Device Class			
Branch	2	3	Total	% of Total
Cardiovascular	16,432	93,041	109,473	31.88%
General Hospital	57,131	3,903	61,034	17.77%
Orthopedic	22,308	21,188	43,496	12.67%
Clinical Chemistry	28,417	5,725	34,142	9.94%
Gastroenterology/Urology	18,020	3,157	21,177	6.17%
Neurology	3,860	13,878	17,738	5.17%
General and Plastic Surgery	16,759	632	17,391	5.06%
Radiology	13,900	6	13,906	4.05%
Ophthalmic	2,295	4,294	6,589	1.92%
Anesthesiology	3,987	52	4,039	1.18%
Hematology	3,686	14	3,700	1.08%
Ear, Nose, and Throat	1,130	2,214	3,344	0.97%
Physical Medicine	2,523	34	2,557	0.74%
Obstetrical and Gynecological	1,610	625	2,235	0.65%
Dental	1,654	76	1,730	0.50%
Toxicology	353		353	0.10%
Immunology	224	29	253	0.07%
Microbiology	131	89	220	0.06%
Pathology	2	13	15	0.00%
Grand Total	194,422	148,970	343,392	

Table 4 - MDR Counts by FDA Branch and Device Class

For this study Cardiovascular and Orthopedic MDR reports were used which provided a potential pool of 152,969 MDR reports. Appendix 3 is a list of all Cardiovascular and Orthopedic product codes with counts by event type of the number of reports for each product code.

As a final step, reports with blank narratives were also excluded resulting in a final candidate pool of 147,727 MDR reports. Table 5 (below) provides a breakdown by month of the candidate pool of MDR reports. From the filtered candidate pool the MDR reports for March, July and October (months 3, 7 and 10) were randomly selected resulting in a document pool with 40,858 documents.

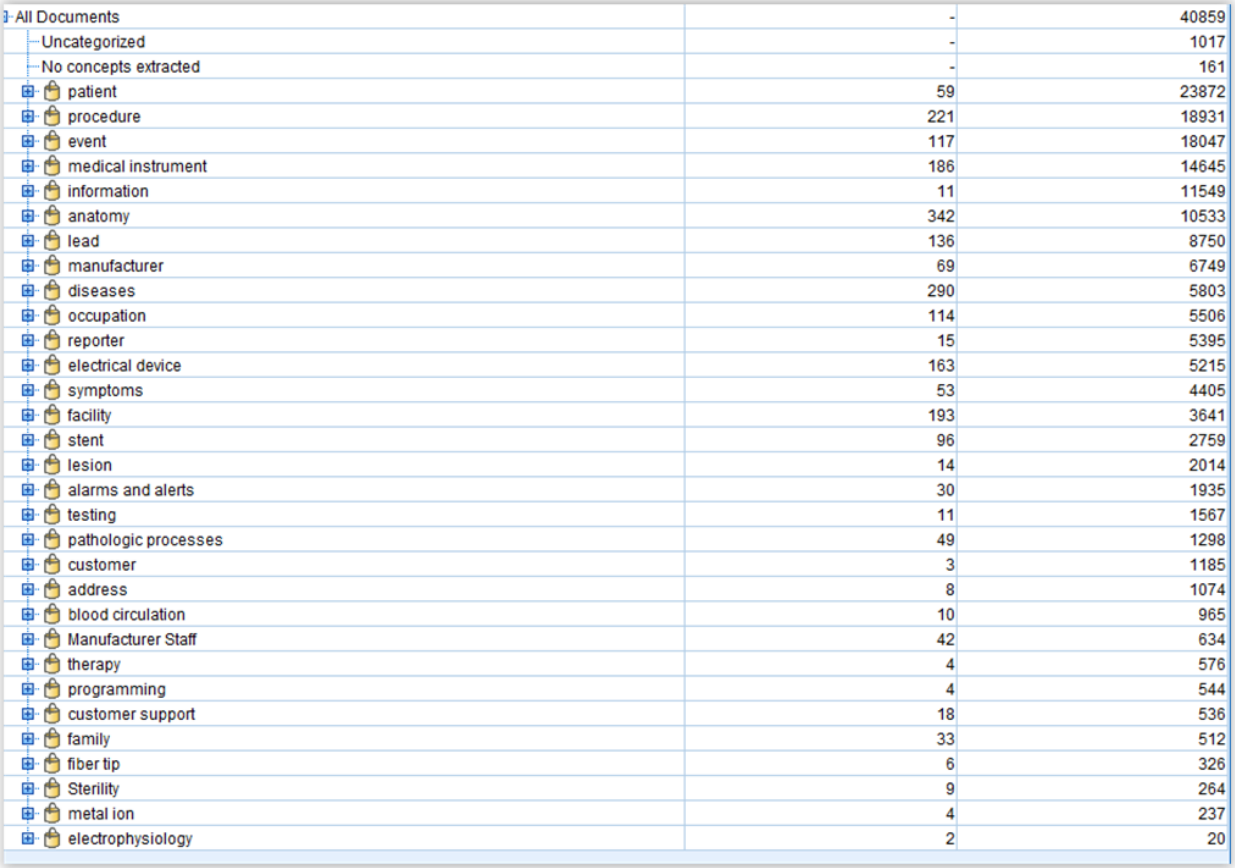
Month	Report Count
1	11,426
2	11,070
3	11,659
4	16,243
5	9,342
6	13,996
7	13,047
8	14,831
9	9,379
10	16,153
11	8,515
12	12,066
Total	147,727

Table 5 - MDR Counts by Month (Filtered Result)

In SPSS, a text mining node was used to extract concepts and create categories. Due to the sublanguage nature of MDR reports, significant effort was devoted to the categorization process as the terms used in this process are not common to the Modeler software. A dictionary of concepts, synonyms and rules was created based on the medical and mechanical nature of the text allowing for better categorization and a more complete view of the data. Many of the default categories created by SPSS were incorrect or misapplied. Various misspellings and different tenses were combined through the use of synonyms. The documents were scored into 31 top level categories with 744 subcategories. For the full model (described below) the count of scored documents by top level categories is found in Figure 2 (below.)

After the text mining, the data was partitioned into testing and training partitions using a 50/50 split. A small number of documents which were initially classified as event type “*” were reclassified as event type “O” (Other.)

A final filter was applied before the modeling phase to remove any categories where the number of results were below 2 as the modeling nodes in SPSS would not be able to process these low volume categories and the information gain for these categories would be minimal.



All Documents	-	40859
Uncategorized	-	1017
No concepts extracted	-	161
patient	59	23872
procedure	221	18931
event	117	18047
medical instrument	186	14645
information	11	11549
anatomy	342	10533
lead	136	8750
manufacturer	69	6749
diseases	290	5803
occupation	114	5506
reporter	15	5395
electrical device	163	5215
symptoms	53	4405
facility	193	3641
stent	96	2759
lesion	14	2014
alarms and alerts	30	1935
testing	11	1567
pathologic processes	49	1298
customer	3	1185
address	8	1074
blood circulation	10	965
Manufacturer Staff	42	634
therapy	4	576
programming	4	544
customer support	18	536
family	33	512
fiber tip	6	326
Sterility	9	264
metal ion	4	237
electrophysiology	2	20

Figure 2- Documents Scored by Category

Models Used

From the data processing steps (see data processing above,) two branches of the FDA were selected: Cardiovascular (Cardio) and Orthopedic (Ortho). Figures 3 and 4 (below) chart the distribution of MDR reports by branch, Table 6 (below) documents report type counts by branch. It is important to note the differences in the data by branch: MDR reports of primary interest, those which will require an investigation, will be heavily skewed to Death (type "D") and Injury (type "I") reports. The frequency of Death reports in the Orthopedic branch is extremely low with the majority of events reported as Injury reports.

The majority of "Death" reports occur in the Cardiology branch, with the remaining reports almost equally split between "Malfunction" and "Injury". These differences by branch represent both an opportunity and a challenge. The opportunity is in creating a unified corpus of terms and concepts which could be applied to all incoming MDR reports. The challenge is in the necessity to compare these results to models utilizing MDR reports from a single FDA branch.

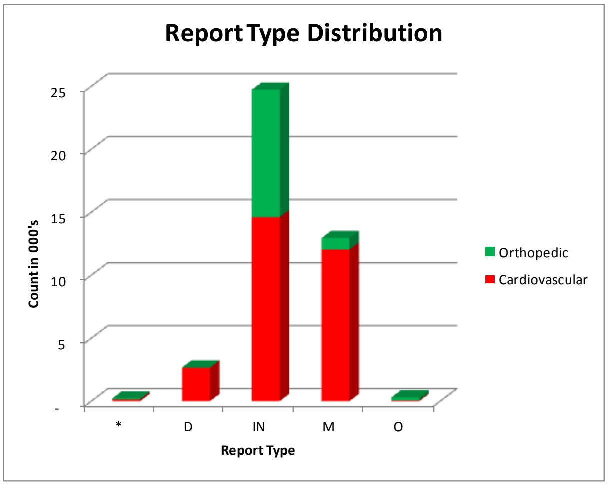


Figure 3 - Event Type Distribution

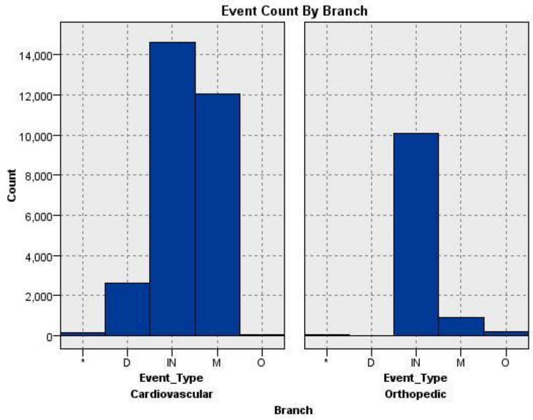


Figure 4 - Histogram of Event Type By Branch

		Branch		
Event Type		Cardiovascular	Orthopedic	Total
*	Count	140	76	216
	Row %	65%	35%	100%
	Column %	0.47%	0.67%	0.53%
D	Count	2,651	21	2,672
	Row %	99%	1%	100%
	Column %	8.98%	0.19%	6.54%
IN	Count	14,607	10,115	24,722
	Row %	59%	41%	100%
	Column %	49.50%	89.13%	60.51%
M	Count	12,038	914	12,952
	Row %	93%	7%	100%
	Column %	40.79%	8.05%	31.70%
O	Count	74	222	296
	Row %	25%	75%	100%
	Column %	0.25%	1.96%	0.72%
Total	Count	29,510	11,348	40,858
	Row %	72%	28%	100%
	Column %	100%	100.00%	100.00%

Table 6 - Report Type

Four models were built, in each model a variety of classification nodes were utilized to predict the report classification. The ‘Full Model’ incorporates the entire set of MDR reports and all categories of concepts. A second model is a variation of the full model utilizing the full data set but only a subset of categories. The final two models split the data by FDA branch but do not change the categorization of concepts.

Full Model

Description

The Full model uses all reports in the pool of data and all categories defined in the text mining node. This model combines MDR reports from both the ‘Orthopedic’ and ‘Cardiovascular’ branches. The Full model initially used an AutoClassifier node in SPSS to determine the best fit models. Once the classification models were found they were run individually and the AutoClassifier node was removed.

Classification nodes run for this model:

- C5
- Logistic Regression
- CHAID
- SVM
- Ensemble (C5, Logistic Regression and CHAID)

Results

The coincidence matrix for the C5 classifier shown in Table 7 (below) indicates an overall accuracy of 82.13%. Given the mix of both Cardiology and Orthopedic reports with different distributions of report types the model appears to be providing reasonably good results.

In the data there are 252 events with the event type “Other”. These reports are of special interest as only 2 of these reports are predicted to be “Other” with the remaining 250 reclassified into the other three main categories. If the “Other” reports are excluded, the C5 classifier accuracy increases to 83.14%.

The distribution of events for the full model shown in Figure 5 and Table 7 (below) shows a skewed distribution with the majority of the events being either “Injury” or “Malfunction”. This is to be expected as one would hope the incidence of “Death” events would be low and the majority of events are classified leaving as few “Other” events as possible. In the full model 1,329 of the 20,556 events (6.47%) are “Death” events, and only 252 (1.23%) are classified as “Other”. The data includes 12,427 (60.45%) “Injury” events and 6,548 (31.85%) “Malfunction” events.

In Table 7 (below,) counts on the diagonal axis represent reports where the predictive model and the actual data have the same event classification. The diagonal axis has 16,882 reports or 82.13% of the total reports.

		Actual				
	C5 Model	D	IN	M	O	Total
Predicted	D (Pred)	990	130	63	4	1,187
	IN (Pred)	216	11,241	1,835	159	13,451
	M (Pred)	123	1,055	4,649	87	5,914
	O (Pred)	-	1	1	2	4
	Total	1,329	12,427	6,548	252	20,556
	% of Total	6.47%	60.45%	31.85%	1.23%	100%

Table 7 – C5 Model Coincidence Matrix

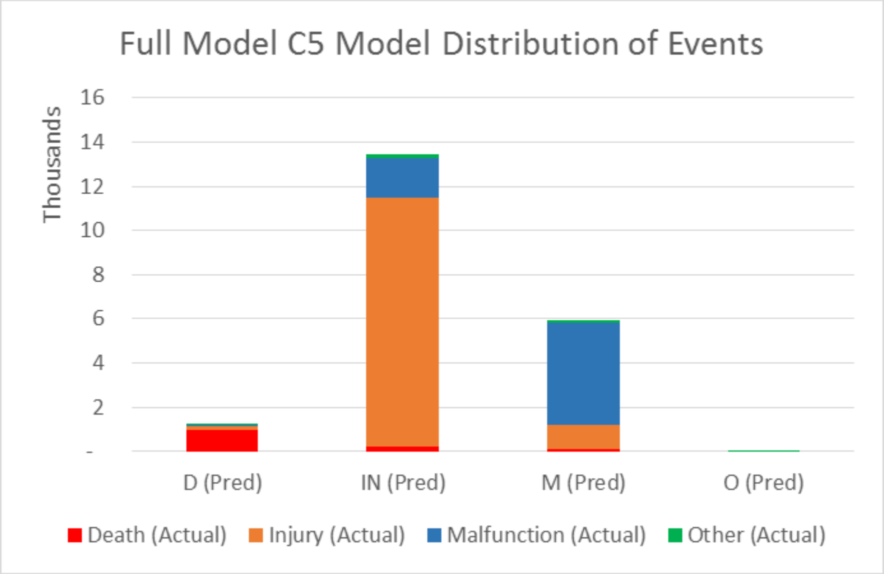


Figure 5 - Full Model C5 Results

The model has predicted a significant number of reports in the “Death” class which does not match the actual data. This could be where the model has identified misclassified reports, or it could be an area for refinement in the model. The model further diverges from the actual data in predicting “Injury” and “Malfunction” moving significant numbers of reports from their original classifications. Here again, this could be a sign significant numbers of reports are misclassified, or an area for further evaluation and refinements to the model to improve the predictive accuracy.

Actual

Predicted

C5 Model		**D**	**IN**	**M**	**O**	**Total**
D	**Count**	990	130	63	4	1,187
	Row %	83.40%	10.95%	5.31%	0.34%	100.00%
	Column %	74.49%	1.05%	0.96%	1.59%	5.77%
IN	**Count**	216	11,241	1,835	159	13,451
	Row %	1.61%	83.57%	13.64%	1.18%	100.00%
	Column %	16.25%	90.46%	28.02%	63.10%	65.44%
M	**Count**	123	1,055	4,649	87	5,914
	Row %	2.08%	17.84%	78.61%	1.47%	100.00%
	Column %	9.26%	8.49%	71.00%	34.52%	28.77%
O	**Count**	0	1	1	2	4
	Row %	0.00%	25.00%	25.00%	50.00%	100.00%
	Column %	0.00%	0.01%	0.02%	0.79%	0.02%
Total	**Count**	1,329	12,427	6,548	252	20,556
	Row %	6.47%	60.45%	31.85%	1.23%	100.00%
	Column %	100.00%	100.00%	100.00%	100.00%	100.00%

Table 8 – Full Model Ensemble Coincidence Matrix

The Ensemble model combines the classification results across the C5, Logistic Regression and CHAID modeling nodes. In the Ensemble model the best result for each grouping is chosen after computing the classification within each modeling node. The Ensemble output produces the coincidence matrix found in Table 8 (above.) The results on the diagonal axis, which show the convergence of the predicted and actual classifications, indicate the Ensemble model is predicting approximately 81.4% (16,727 out of 20,556) of the reports to match the actual data. While this is encouraging, the model is still predicting a significant number of events which do not match the classification found in the data. As we are assuming some reports may be classified incorrectly in the actual data, a review of the non-matching data is necessary to determine the accuracy of the model.

The Ensemble model predicted 78 events of type ‘Other’, which is fewer than the 252 actual ‘Other’ events, but still significant. For the model to be successful the ‘Other’ type events must be minimized. The coincidence matrix results for this model can be found in Appendix 4.

Select Model

Description

The Select model uses the same source data as the Full model with no filtering or changes. The Select model attempts to improve on the Full model results by utilizing a grouping of categories from the text analytics node based on the results of the C5 and CHAID classification in the Full Model. Classification nodes run for this model:

- C5
- Logistic Regression
- CHAID
- SVM
- Ensemble (C5, Logistic Regression and CHAID)

Results

The C5 model coincidence matrix in Table 9 (below) indicates an overall model accuracy of 76.65% (15,756 reports out of 20,556 on the diagonal axis.) If the “Other” reports are excluded, the model the accuracy is 77.58%. A significant number of events are predicted as “Injury” events including 43% (575 out of 1,329) of the “Death” reports. While only a manual review of these reports which have been ‘downgraded’ from “Death” to “Injury” can validate the model results, the large concentration of predicted results as “Injury” (15,302 out of 20,556 or almost 75%) is concerning. The distribution of report classification should be similar to the original data, but the Select Model displays a heavy concentration of results in just a single class.

Because the source data is the same as the Full model, the distribution of actual report types in this model will be exactly the same as the Full Model. A similar concern can be raised from the coincidence matrix of the Ensemble classification where the concentration of predicted “Injury” reports is even higher at 15,720 or just over 76%. (See Table 10 below.)

Actual

Predicted

C5 Model	D	IN	M	O	Total
D	685	77	38	1	801
IN	575	11,555	3,004	168	15,302
M	69	786	3,505	72	4,432
O	-	9	1	11	21
Total	1,329	12,427	6,548	252	20,556
% of Total	6.47%	60.45%	31.85%	1.23%	100%

Table 9 – C5 Model Coincidence Matrix

Actual

Predicted

Ensemble Model		D	IN	M	O	Total
D	Count	690	75	66	1	832
	Row %	82.93%	9.01%	7.93%	0.12%	100.00%
	Column %	51.92%	0.60%	1.01%	0.40%	4.05%
IN	Count	572	11,734	3,245	169	15,720
	Row %	3.64%	74.64%	20.64%	1.08%	100.00%
	Column %	43.04%	94.42%	49.56%	67.06%	76.47%
M	Count	65	612	3,234	74	3,985
	Row %	1.63%	15.36%	81.15%	1.86%	100.00%
	Column %	4.89%	4.92%	49.39%	29.37%	19.39%
O	Count	2	6	3	8	19
	Row %	10.53%	31.58%	15.79%	42.11%	100.00%
	Column %	0.15%	0.05%	0.05%	3.17%	0.09%
Total	Count	1,329	12,427	6,548	252	20,556
	Row %	6.47%	60.45%	31.85%	1.23%	100.00%
	Column %	100.00%	100.00%	100.00%	100.00%	100.00%

Table 10 - Select Model Average Coincidence Matrix

The Ensemble model results for the Select Model in Table 10 (above) indicate the model may have a problem predicting the report class. The number of predicted “Death” reports is only 62% of the actual data, and “Injury” reports are 126% of the actual data leading to the possible conclusion the model requires additional refinement. The Ensemble model has an overall accuracy of 76.21% (15,666 out of 20,556 reports) which is lower than the full model. While the select model did a better job of reducing the “Other” predictions with only 19 vs. 78 in the full model, the model is predicting a significant number of “Malfunction” reports as “Injury” (3,245 reports.)

The Ensemble model classified 19 reports as “Other”, which improves on the 252 found in the data. Of the 19 classified by the model as “Other”, 8 are also classified as “Other” in data which indicates the model may be better at dealing with poorly classified reports, but the overall accuracy of 76% is disappointing as it is less than the full model. The coincidence matrix results for this model can be found in Appendix 4.

Cardio Model

Description

The Cardio model incorporates data for the same three months (March, July, and October) as the previous models, with a filter applied to limit the model to only those reports sent to the FDA Cardiology branch. The text mining and classification nodes were run and used results specific to the text mining categories found in the Cardiology branch. No changes were made to the concepts or the resulting categories, and there was no filtering of the categories – all categories were utilized as they were output from the text mining node.

The incoming data will include the majority of the “Death” reports, but otherwise resembles the distribution seen in with the Full model: Approximately an even split between the “Injury” and “Malfunction” reports, with a slight skew towards “Injury”. The Cardio model ran with 14,877 reports. The Logistic Regression node did not complete and was excluded from the model. Classification nodes run for this model:

- C5
- CHAID
- SVM
- Ensemble (C5, CHAID and SVM)

Results

The model predicted a distribution of report classes similar to the incoming data. The model appears to benefit from the separation of the incoming reports by FDA branch (where in the previous models Cardio and Ortho reports were combined which may have had a dampening effect on model accuracy.) The coincidence matrix for the C5 model found in Table 11 (below) indicates an accuracy of 80% (11,911 out of 14,877 reports.) This value increases only marginally to 80.5% if the "Other" reports are excluded.

The model is predicting 78% of the "Death" reports, but the remaining 22% (293 reports) are an area of concern. As with the previous models these reports would require a manual review to ensure the accuracy of the prediction. The model is also predicting an additional 127 reports to be "Death" reports, reclassifying the reports from other classes. A significant group of reports have been classified as "Malfunction" from "Injury". This may be an indication of a problem in the categorization in the text mining node.

		Actual				
	C5 Model	D	IN	M	O	Total
Predicted	D (Pred)	1,042	73	54	-	1,169
	IN (Pred)	207	6,274	1,404	23	7,908
	M (Pred)	85	1,055	4,566	45	5,751
	O (Pred)	1	6	13	29	49
	Total	1,335	7,408	6,037	97	14,877
	% of Total	8.97%	49.79%	40.58%	0.65%	100%

Table 11 - C5 Model Coincidence Matrix

The incoming data has a higher concentration of "Death" reports with approximately 9% of all reports classified as "Death" and nearly 50% of all reports classified as "Injury". This distribution fits with the expectation that Cardiology adverse events would have a higher likelihood of causing either a death or an injury. Comparing a graph of the C5 model results from the full model in Figure 5 (repeated below) to the C5 results of the Cardio model in Figure 6 (below) shows the result of the full model are more centered on the "Injury" classification where the Cardio model, while still skewed, has a flatter distribution with just under 50% of the reports classified as "Injury" and another 40.6% classified as "Malfunction".

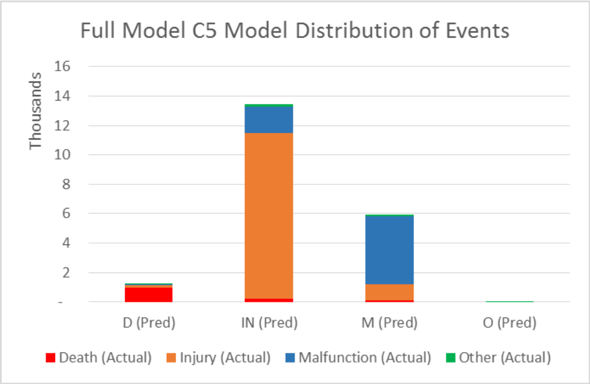


Figure 6 - Full Model C5 Results

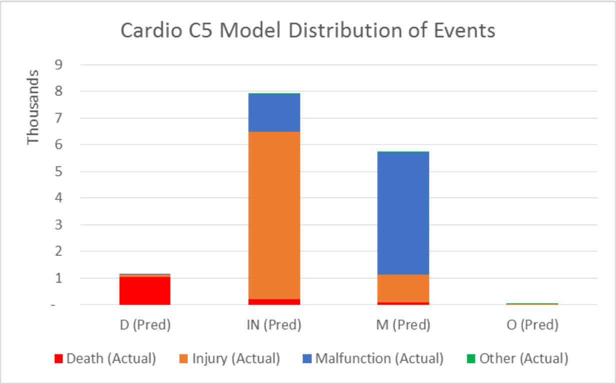


Figure 7 – Cardio Model C5 Results

The Ensemble model coincidence matrix is found in Table 12 (below) with the diagonal axis representing 82.15% (12,222 out of 14,877) of the reports. The model appears to be performing well in finding "Death" reports with 94% of the predicted "Death" reports agreeing with the classification in the data. The Ensemble model doesn't predict "Injury" or "Malfunction" reports as well with accuracy rates of 80% and 82.5% respectively. The Ensemble model is only predicting 42 "Other" reports, an improvement over the 97 found in the data.

Actual

Predicted

Ensemble Model		**D**	**IN**	**M**	**O**	**Total**
D	**Count**	1,018	29	34	0	1,081
	Row %	94.17%	2.68%	3.15%	0.00%	100.00%
	Column %	76.25%	0.39%	0.56%	0.00%	7.27%
IN	**Count**	226	6,531	1,349	20	8,126
	Row %	2.78%	80.37%	16.60%	0.25%	100.00%
	Column %	16.93%	88.16%	22.35%	20.62%	54.62%
M	**Count**	90	846	4,644	48	5,628
	Row %	1.60%	15.03%	82.52%	0.85%	100.00%
	Column %	6.74%	11.42%	76.93%	49.48%	37.83%
O	**Count**	1	2	10	29	42
	Row %	2.38%	4.76%	23.81%	69.05%	100.00%
	Column %	0.07%	0.03%	0.17%	29.90%	0.28%
Total	**Count**	1,335	7,408	6,037	97	14,877
	Row %	8.97%	49.79%	40.58%	0.65%	100.00%
	Column %	100.00%	100.00%	100.00%	100.00%	100.00%

Table 12 - Cardio Model Average Coincidence Matrix

The coincidence matrix results for this model can be found in Appendix 4.

Ortho Model

Description

The Ortho model incorporates data for the same three months (March, July, and October) with a filter applied to limit the model to only those reports sent to the FDA Orthopedic branch. The text mining and classification nodes were run and used results specific to the text mining categories found in the Orthopedic branch. No changes were made to the concepts or the resulting categories, and there was no filtering of the categories – all categories were utilized as they were output from the text mining node.

The incoming data included very few (12) "Death" reports, and is heavily skewed towards the "Injury" classification with 89% of all reports in this group. This matches expectations as an Orthopedic product is more likely to cause an injury than a death. The heavy concentration of "Injury" reports may be an indication of good classification of reports in this branch with few errors. The Logistic Regression node did not complete and was excluded from the model. Classification nodes run for this model:

- C5
- CHAID
- SVM
- Ensemble (C5, CHAID and SVM)

Results

The Orthopedic model classified 5,721 MDR reports which were heavily concentrated in the "Injury" classification (89%) see Table 13 (below.) With only 12 "Death" reports predicted the model may not have had enough data to correctly classify incoming reports into the "Death" category. The majority of the reports in both the actual data and the predicted classification remain in the "Injury" classification making this model somewhat one dimensional. The 5,222 reports found on the diagonal axis give the model a 91.28% accuracy rate which may be misleading.

Actual

Predicted

C5 Model	D	IN	M	O	Total
D (Pred)	-	1	-	-	1
IN (Pred)	11	5,000	259	109	5,379
M (Pred)	1	95	197	9	302
O (Pred)	-	13	1	25	39
Total	12	5,109	457	143	5,721
% of Total	0.21%	89.30%	7.99%	2.50%	100.00%

Table 13 - C5 Model Coincidence Matrix

The singular dimensionality of the data is easily seen in Figure 7 (below) with the “Injury” classification dwarfing all other groupings. As previously mentioned, this concentration of reports into a single class may be due to good classification of the initial reports, but does not give the model much data to work with in other report classes.

The Ensemble model coincidence matrix is found in Table 14 (below.) The Ensemble model classified 92% of the reports in agreement with the classification found in the data (5,271 out of 5,721.) The model predicted no “Death” reports moving the 12 found in the data to “Injury” reports. The model also reduced the number of “Other” reports from 143 to only 22 with 21 of these 22 matching the classifications found in the data.

The Ensemble model overall classified 95.42% of the reports as “Injury” which is higher than the 89.3% classified in the data. The majority of the changes are found in “Malfunction” and “Other” reports which have been re-classified as “Injury”.

The coincidence matrix results for this model can be found in Appendix 4.

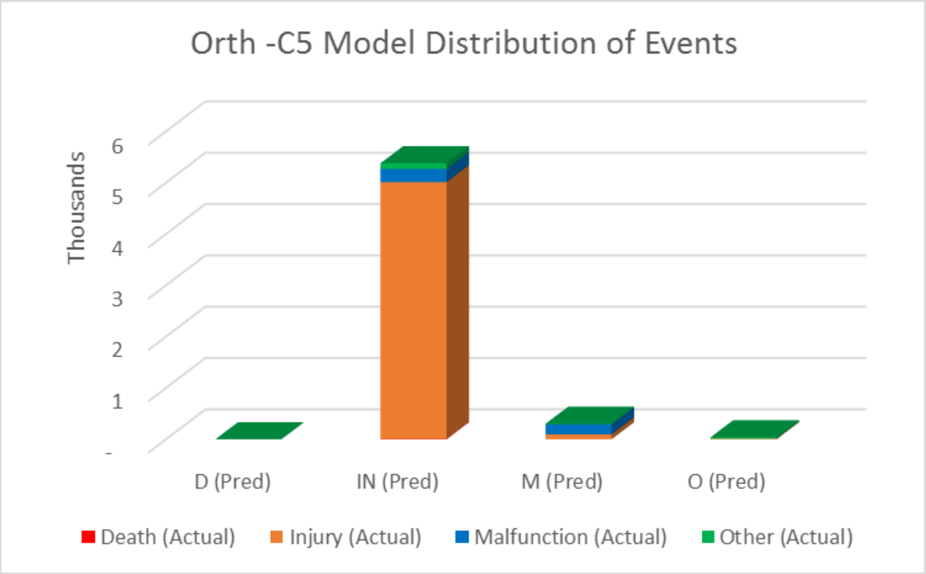


Figure 8 - C5 Results

Actual

Predicted

Ensemble Model		**D**	**IN**	**M**	**O**	**Total**
D	**Count**	0	0	0	0	0
	Row %	0.00%	0.00%	0.00%	0.00%	0.00%
	Column %	0.00%	0.00%	0.00%	0.00%	0.00%
IN	**Count**	12	5,063	270	114	5,459
	Row %	0.22%	92.75%	4.95%	2.09%	100.00%
	Column %	100.00%	99.10%	59.08%	79.72%	95.42%
M	**Count**	0	45	187	8	240
	Row %	0.00%	18.75%	77.92%	3.33%	100.00%
	Column %	0.00%	0.88%	40.92%	5.59%	4.20%
O	**Count**	0	1	0	21	22
	Row %	0.00%	4.55%	0.00%	95.45%	100.00%
	Column %	0.00%	0.02%	0.00%	14.69%	0.38%
Total	**Count**	12	5,109	457	143	5,721
	Row %	0.21%	89.30%	7.99%	2.50%	100.00%
	Column %	100.00%	100.00%	100.00%	100.00%	100.00%

Table 14 - Ortho Model Ensemble Coincidence Matrix

Summary

The intent of the work is to determine if Text Analytics can be used to classify medical device adverse event reports. The automation of the classification as the reports are received would allow medical device manufacturers a better method for triaging incoming reports and the decision process of which reports to investigate and report back to the FDA. The triaging of incoming reports is currently a largely manual task. The automation would benefit the manufacturer by confirming those reports which either do or do not require an investigation, and also by identifying misclassified reports which, in fact, also require investigation.

To accomplish the Text Analysis task three months of data has been analyzed for two groups of medical device product codes which combined represent product codes reviewed by the Cardiology and Orthopedic branches of the FDA.

Significant effort devoted to proper categorization of the concepts found in the data as the report sublanguage presents a unique mix of medical terminology and sentence structure. While it was not a goal of the categorization to minimize the number of categories, it was understood the model accuracy could suffer if the number of categories was high, especially if the categories added little or no information to the model. Correctly grouping the concepts into types and categories thus was crucial to the overall process. Poorly grouped concepts would result in poor results. Extensive use of text rules and synonyms helped in the categorization of concepts.

Once the concept categorization was complete the data was split 50/50 into test and training datasets. This allowed the models to 'learn' as the data was processed. Initially an Auto-classifier node was used determine which classification nodes would run to completion and which classification nodes would work best with the data. Once the classification nodes were generated from the auto-classifier these nodes were used exclusively.

Four models were built, each with multiple classification nodes. The models were:

- Full model – Use all categories from the text mining node with no changes
- Select model – Identify a subset of concept categories based on the results of the Full model, use just these categories. For this model the individual modeling nodes were re-run to ensure correct results from the subset of categories.
- Cardio model – Create a subset of data based on the group of Product Codes reviewed by the Cardiology branch of the FDA. For this model text mining and individual classification nodes were instantiated. The reports in this model contained the majority of the "Death" reports. "Injury" and "Malfunction" reports were almost evenly split.
- Ortho model – Create a subset of data based on the group of Product Codes reviewed by the Orthopedic branch of the FDA. For this model text mining and individual classification nodes were instantiated. The MDR reports in this model contained virtually no "Death" events and were dominated by "Injury" events.

The classification nodes in common between the models were:

- C5
- CHAID
- SVM
- Ensemble

A logistic regression classification node was used in some models but did not complete in all models.

Conclusions

Table 15 (below) is a summary of the results from the various models and methods of classification. From these results it is clear the Select model did poorly when compared to the other models. The Full model, which represents a combination of concept categorizations across both the Cardiology and Orthopedic branches, results are similar to the results of the Cardio model. This may be due to similar distributions of report types between the Full and Cardio models. The report type distribution for the Ortho model is significantly different from the Full model with the reports in the Ortho model heavily concentrated in a single report type. Reports to the Ortho and Cardio branches appear to have very different data characteristics which may have had a dampening effect on the Full model results. The Ortho model does have strong results, but these results may be at least in part due to the heavy bias in the data towards "Injury" reports.

	Classification Method			
Model	C5	CHAID	SVM	Ensemble
Full	82.13%	79.12%	83.17%	81.37%
Select	76.65%	68.73%	72.32%	76.21%
Cardio	80.03%	75.19%	82.18%	82.15%
Ortho	91.28%	90.58%	92.38%	92.13%

Table 15 - Summary of Model Results

The differences in the classification of the MDR reports between the Cardio and Ortho data was a major consideration in splitting the data and running the models independently. The Ortho and Cardio models were run with no revisions made to the concept categorization allowing them to utilize the same concepts and structures defined for the Full model. From the results it would seem this logic is sound as the results are at least as good as the Full model and, for the Ortho model in particular, even better.

Splitting the data by FDA branch is a better reflection of reality. If the models were deployed to a manufacturer, it is highly unlikely the manufacturer would need to triage incoming reports from both FDA branches as manufacturers tend to remain focused in a core set of device types.

As such it is very possible the branch specific models make better sense and could be utilized by a manufacturer.

Within the models the SVM classification provided the highest model accuracy in three out of four models with the ensemble method also producing strong results. Both of these observations should not be considered a surprise as the SVM classification will tend to provide strong results and the ensemble method will lean towards the SVM results whenever the SVM classification provides the best alternative.

Overall the work here shows promise in the prediction of report class based on event narrative. Creation of individual term libraries by FDA branch and further refinement of those concepts and categories would likely result in deployable models. Once deployed a model classifying incoming reports would allow the device manufacturer to automate the triage of incoming data. This would simplify the process by allowing the manufacturer to focus on only those reports which actually require review and investigation where today all reports are manually reviewed and assessed. The automation would also provide the manufacturer with a method of identifying the group of reports which were misclassified in the data reducing the opportunity to overlook reports requiring action. A further benefit of an automated solution is the reduction in the number of “Other” type reports requiring review.

Overall the automations’ goal is to improve the accuracy and the efficiency of the incoming report triage process. The automation has, while it is not its’ goal, the possibility of saving the manufacturer money by automating the triage process.

Further Work and Revisions

This work shows potential for future application in a number of areas. Before the models could be deployed addition refinement of the concepts and categorization would be required with an eye towards improving the model results. Further refinement of branch specific terminology should be considered and a set of libraries with the branch specific concepts and text rules created. A set of rules would also be necessary for the occasion where a model is run with data from more than one branch.

This work could also be extended easily into other ticket or log type environments such as help desk tickets, maintenance logs or other work logs. While each of these environments utilize a unique sublanguage, the methods and concepts used in this work could easily be extended to those areas.

It is difficult to ascertain the true accuracy of the various models as one of the foundational assumptions for this work has been that at least a portion of the reports were not classified correctly. As such purely reviewing the diagonal axis of the coincidence matrix will not necessarily yield the complete answer to the question of model accuracy. It will only yield the proportion of reports where the predicted classification matches the classification in the data. To validate the accuracy of the models a manual review of the data, especially where the prediction and actual classification diverge, is required.

In many (but not all) of the works reviewed for this paper the authors were forced to use relatively small data sets as the amount of data available was not large. In this work there are many years of data freely available for study. The enormous volume of data available allows the models to be run in a variety of combinations of years and months resulting in a wide range of possible quantities of MDR reports. The readily available data would also allow for additional testing and refinement of the models with an eye towards maximizing the overall model accuracy. In the future testing a set of models with data combined across FDA branches versus models running on data for one specific branch at a time should be considered.

The data used in this study was drawn from three randomly selected months from one randomly selected year. Consideration should also be given to expanding the initial data to capture concepts in a larger population of data by either utilizing more months, years or both.

Appendix 1 – Top 50 Product Codes by MDR Count

(Cardiology and Orthopedic branches with > 1.5% “Death” or “”Injury” reports highlighted in green)

Product Code	Event						Grand Total	% of Total	% Death / Injury	Branch	Device Type
	*	D	IN	M	O	(blank)					
FRN	57	41	429	24,061	33		24,621	9.05%	0.17%	HO	Pump, Infusion
NBW	24	13	2,986	18,945	7		21,975	8.08%	1.10%	CH	System, Test, Blood Glucose, Over The Counter
KWA	21	35	17,182	246	122	1	17,607	6.47%	6.33%	OR	Prosthesis, Hip, Semi-Constrained (Metal Uncemented Acetabular Component)
LWS	9	2,187	7,586	5,213	13		15,008	5.52%	3.59%	CV	Implantable Cardioverter Defibrillator (Non-Crt)
DTB		1,769	7,047	3,659	1		12,476	4.59%	3.24%	CV	Permanent Pacemaker Electrode
JAA	5	3	11	12,318			12,337	4.54%	0.01%	RA	System, X-Ray, Fluoroscopic, Image-Intensified
LGW	11	41	7,361	3,325	49		10,787	3.97%	2.72%	NE	Stimulator, Spinal-Cord, Totally Implanted For Pain Relief
FNL	63	25	62	9,995	69		10,214	3.76%	0.03%	HO	Bed, Ac-Powered Adjustable Hospital
LZG	27	79	3,039	6,145	21		9,311	3.42%	1.15%	HO	Pump, Infusion, Insulin
NIQ	20	761	4,361	1,945	1		7,088	2.61%	1.88%	CV	Coronary Drug-Eluting Stent
MKJ	20	221	26	6,351	13		6,631	2.44%	0.09%	CV	Automated External Defibrillators (Non-Wearable)
NVN	1	151	4,580	1,481			6,213	2.28%	1.74%	CV	Drug Eluting Permanent Right Ventricular (Rv) Or Right Atrial (Ra) Pacemaker Electrodes
NVZ	2	507	3,119	2,429	58		6,115	2.25%	1.33%	CV	Pulse Generator, Permanent, Implantable
FKX	8	91	1,201	4,781			6,081	2.24%	0.48%	GU	System, Peritoneal, Automatic Delivery
NVY		125	4,643	1,262			6,030	2.22%	1.75%	CV	Permanent Defibrillator Electrodes
DXY		181	4,188	1,636	3		6,008	2.21%	1.61%	CV	Implantable Pacemaker Pulse-Generator
NIK	1	500	3,795	1,673			5,969	2.20%	1.58%	CV	Defibrillator, Automatic Implantable Cardioverter, With Cardiac Resynchronization (Crt-D)
MDS	8	84	4,019	1,483	130		5,724	2.10%	1.51%	CH	Sensor, Glucose, Invasive
JWH	9	3	4,516	304	158		4,990	1.84%	1.66%	OR	Prosthesis, Knee, Patellofemorotibial, Semi-Constrained, Cemented, Polymer/Metal/Polymer
FPO	7	1	62	4,504	12		4,586	1.69%	0.02%	HO	Stretcher, Wheeled
KDJ	11	56	3,476	365			3,908	1.44%	1.30%	GU	Set, Administration, For Peritoneal Dialysis, Disposable
LKK	3	46	2,618	1,227			3,894	1.43%	0.98%	HO	Pump, Infusion, Implanted, Programmable
MGB	3	37	3,734	108	1		3,883	1.43%	1.39%	CV	Device, Hemostasis, Vascular
LFR			142	3,660	3		3,805	1.40%	0.05%	CH	Glucose Dehydrogenase, Glucose
FTL	10	24	3,226	252	32	7	3,551	1.31%	1.20%	SU	Mesh, Surgical, Polymeric

LPH	9	3	3,265	96	63		3,436	1.26%	1.20%	OR	Prosthesis, Hip, Semi-Constrained, Metal/Polymer, Porous Uncemented
MIH	4	448	2,327	304	3		3,086	1.13%	1.02%	CV	System, Endovascular Graft, Aortic Aneurysm Treatment
OJX	1	96	1,924	1,017	1		3,039	1.12%	0.74%	CV	Drug Eluting Permanent Left Ventricular (Lv) Pacemaker Electrode
GDW	17	47	1,403	1,566	3		3,036	1.12%	0.53%	SU	Staple, Implantable
JDI	8	5	2,820	131	35		2,999	1.10%	1.04%	OR	Prosthesis, Hip, Semi-Constrained, Metal/Polymer, Cemented
FPA	24	33	268	2,444	19		2,788	1.03%	0.11%	HO	Set, Administration, Intravascular
GEX	9		174	2,482	10		2,675	0.98%	0.06%	SU	Powered Laser Surgical Instrument
GJS	9	8	295	2,235	101		2,648	0.97%	0.11%	HE	Test, Time, Prothrombin
HQL	20		1,557	664	316		2,557	0.94%	0.57%	OP	Intraocular Lens
GEI	65	16	663	1,692	41		2,477	0.91%	0.25%	SU	Electrosurgical, Cutting & Coagulation & Accessories
MCM			864	1,273	2		2,139	0.79%	0.32%	EN	Implant, Cochlear
LOX	7	22	265	1,783			2,077	0.76%	0.11%	CV	Catheters, Transluminal Coronary Angioplasty, Percutaneous
DRY	191			1,751	15		1,957	0.72%	0.00%	CV	Monitor, Blood-Gas, On-Line, Cardiopulmonary Bypass
DTE		5	108	1,748			1,861	0.68%	0.04%	CV	Pulse-Generator, Pacemaker, External
KDI	6	32	1,299	433	3		1,773	0.65%	0.49%	GU	Dialyzer, High Permeability With Or Without Sealed Dialysate System
MEB	2	7	313	1,395	15		1,732	0.64%	0.12%	HO	Pump, Infusion, Elastomeric
EZW	2	3	700	890	1		1,596	0.59%	0.26%	GU	Stimulator, Electrical, Implantable, For Incontinence
MHY	5	21	676	849	1		1,552	0.57%	0.26%	NE	Stimulator, Electrical, Implanted, For Parkinsonian Tremor
HQC	2		395	1,087	46		1,530	0.56%	0.15%	OP	Unit, Phacofragmentation
GCJ	120	11	143	1,183	11		1,468	0.54%	0.06%	SU	Laparoscope, General & Plastic Surgery
INK	4	3	40	1,383	21		1,451	0.53%	0.02%	PM	Stretcher, Wheeled, Powered
MVK	1	22	1	1,371			1,395	0.51%	0.01%	CV	Wearable Automated External Defibrillator
LYJ	2	158	1,108	67			1,335	0.49%	0.47%	NE	Stimulator, Autonomic Nerve, Implanted For Epilepsy
FGE	4	13	456	799	2		1,274	0.47%	0.17%	GU	Catheter, Biliary, Diagnostic
MMI	10		324	903	4		1,241	0.46%	0.12%	CH	Immunoassay Method, Troponin Subunit
Grand Total	842	7,934	114,797	146,914	1,439	8	271,934				

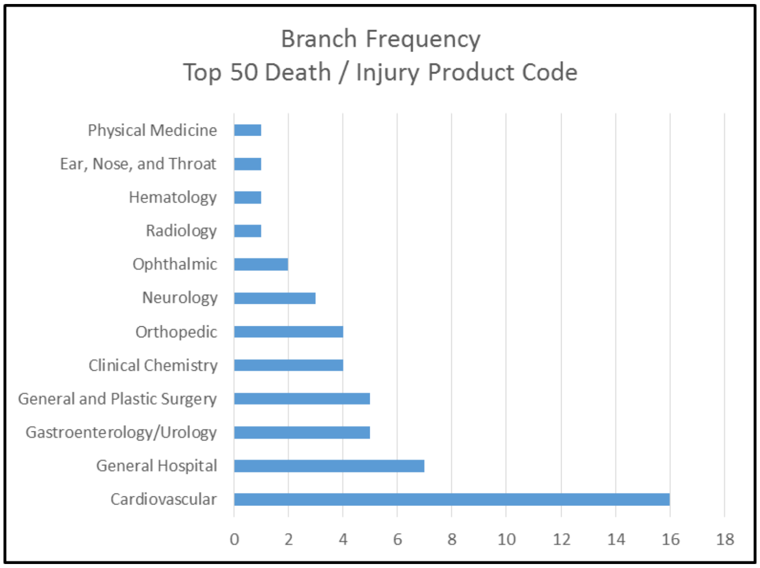
Appendix 2 – MDR Data Analysis

Product Codes with at least 1.5% Death and Injury Reports

Product Code	Total Reports	% Death / Injury	Branch
KWA	17,607	6.33%	Orthopedic
LWS	15,008	3.59%	Cardiovascular
DTB	12,476	3.24%	Cardiovascular
LGW	10,787	2.72%	Neurology
NIQ	7,088	1.88%	Cardiovascular
NVN	6,213	1.74%	Cardiovascular
NVY	6,030	1.75%	Cardiovascular
DXY	6,008	1.61%	Cardiovascular
NIK	5,969	1.58%	Cardiovascular
MDS	5,724	1.51%	Clinical Chemistry
JWH	4,990	1.66%	Orthopedic

Branch Frequency – Death / Injury Top 50 Product Codes

Branch	Name	Count
RA	Radiology	1
HE	Hematology	1
EN	Ear, Nose, and Throat	1
PM	Physical Medicine	1
OP	Ophthalmic	2
NE	Neurology	3
CH	Clinical Chemistry	4
OR	Orthopedic	4
GU	Gastroenterology/Urology	5
SU	General and Plastic Surgery	5
HO	General Hospital	7
CV	Cardiovascular	16



Appendix 3 – Cardiovascular / Orthopedic Product Code and Event Types

Cardiovascular Product Codes

	Event						
Product Code	*	D	IN	M	O	(blank)	Grand Total
Cardiovascular							
LWS	9	2,187	7,586	5,213	13		15,008
DTB		1,769	7,047	3,659	1		12,476
NIQ	20	761	4,361	1,945	1		7,088
MKJ	20	221	26	6,351	13		6,631
NVN	1	151	4,580	1,481			6,213
NVZ	2	507	3,119	2,429	58		6,115
NVY		125	4,643	1,262			6,030
DXY		181	4,188	1,636	3		6,008
NIK	1	500	3,795	1,673			5,969
MGB	3	37	3,734	108	1		3,883
MIH	4	448	2,327	304	3		3,086
OJX	1	96	1,924	1,017	1		3,039
LOX	7	22	265	1,783			2,077
DRY	191			1,751	15		1,957
DTE		5	108	1,748			1,861
MVK	1	22	1	1,371			1,395
MAF	3	98	531	596			1,228
DQX	7	34	478	706	2		1,227
DYE	13	191	990	24	3		1,221
DSQ	4	94	763	161			1,022
DQY	3	32	296	658	6		995
DTQ		9	5	856	25		895
LIT	4	3	165	637	5		814
MHX	13	121	55	574	2		765
NIM		97	552	72			721
KRG		1	9	704			714
DSI	1	18	193	458			670
DYB	7	12	290	195			504
DQO	3	7	146	344			500
DSP	6	28	49	397	1		481
LWP		29	283	148	1		461
MRM		31	174	207	2		414
OAD	3	21	191	179			394
MCX		29	117	183	1		330
DTK	5	7	208	103			323
KRH		85	230	4	1		320
LDD	1	3	4	307	3	1	319
NTE	1	5	208	84			298
DWC				249			249
NIO		3	151	94			248
LWR	1	26	207	8	1		243
DYG	4	3	54	174	1		236
JOR				217			217
DRF	4	2	115	87			208

KRA	2	11	29	153	1		196
DWF	7	11	33	140	4		195
MCW	2	1	171	20			194
NKE		6	126	53	7		192
MLV	8	3	170	3			184
OEZ	14	2	4	150	2		172
Cardiovascular Total	376	8,055	54,701	42,676	177	1	105,986

Orthopedic Product Codes

	Event						
Product Code	*	D	IN	M	O	(blank)	Grand Total
Orthopedic							
KWA	21	35	17,182	246	122	1	17,607
JWH	9	3	4,516	304	158		4,990
LPH	9	3	3,265	96	63		3,436
JDI	8	5	2,820	131	35		2,999
NJL			1,087	11			1,098
HWC	8	1	725	227	41		1,002
KWY	2	2	723	44	16	2	789
KWP	26	4	467	244	1		742
HRS	17	1	545	123	8		694
KXA	1		605	11	2	4	623
HSB	4	1	447	116			568
LZO	5		470	32	9		516
HRX	7	9	108	372	7		503
NKB	4		390	89	2		485
KWS			249	30	162		441
NXT	1		420	5			426
LOD	1	8	313	70	4	2	398
MEH	1		344	40	2	2	389
KWQ	7	2	215	122	2		348
NEK	1	10	301	1			313
MBI			121	145	22		288
JDL			266	6			272
LWJ		2	244	5	3		254
MRA	1		223	19			243
HSD			201	26	7		234
MJO	7		183	37			227
MBH	1		166	47			214
MOZ	57	1	113		40		211
KTT	1	7	143	26			177
NDJ			171				171
MAX	7	1	66	93	2		169
HRY			148	4	1		153
NRA			141	2			143
HSN			136	2			138
MAI	2		83	38			123
HTG			115	2	6		123
MNI	4		72	44	2		122
KWZ	1	1	97	13	3		115

MQP	3		52	53			108
KRO	3		84	15			102
NDN		6	32	58			96
HSX			76	19			95
KWT			79	12	4		95
JDS			82	11			93
MQV	1	1	72	8	7		89
LXT			54	29			83
MNH			65	16			81
KYI			71				71
JDC			56	5			61
NQP	1		48	7		1	57
Orthopedic Total	221	103	38,652	3,056	731	12	42,775

Appendix 4 – Model Output

Full Model

Actual

Predicted

C5 Model		D	IN	M	O	Total
D	Count	990	130	63	4	1,187
	Row %	83.40%	10.95%	5.31%	0.34%	100.00%
	Column %	74.49%	1.05%	0.96%	1.59%	5.77%
IN	Count	216	11,241	1,835	159	13,451
	Row %	1.61%	83.57%	13.64%	1.18%	100.00%
	Column %	16.25%	90.46%	28.02%	63.10%	65.44%
M	Count	123	1,055	4,649	87	5,914
	Row %	2.08%	17.84%	78.61%	1.47%	100.00%
	Column %	9.26%	8.49%	71.00%	34.52%	28.77%
O	Count	0	1	1	2	4
	Row %	0.00%	25.00%	25.00%	50.00%	100.00%
	Column %	0.00%	0.01%	0.02%	0.79%	0.02%
Total	Count	1,329	12,427	6,548	252	20,556
	Row %	6.47%	60.45%	31.85%	1.23%	100.00%
	Column %	100.00%	100.00%	100.00%	100.00%	100.00%

Actual

Predicted

CHAID Model		D	IN	M	O	Total
D	Count	967	205	186	7	1,365
	Row %	70.84%	15.02%	13.63%	0.51%	100.00%
	Column %	72.76%	1.65%	2.84%	2.78%	6.64%
IN	Count	261	11,519	2,584	175	14,539
	Row %	1.80%	79.23%	17.77%	1.20%	100.00%
	Column %	19.64%	92.69%	39.46%	69.44%	70.73%
M	Count	101	703	3,778	70	4,652
	Row %	2.17%	15.11%	81.21%	1.50%	100.00%
	Column %	7.60%	5.66%	57.70%	27.78%	22.63%
Total	Count	1,329	12,427	6,548	252	20,556
	Row %	6.47%	60.45%	31.85%	1.23%	100.00%
	Column %	100.00%	100.00%	100.00%	100.00%	100.00%

Actual

Predicted

Logistic Regession Model		D	IN	M	O	Total
Null	Count	3	16	13	5	37
	Row %	8.11%	43.24%	35.14%	13.51%	100.00%
	Column %	0.23%	0.13%	0.20%	1.98%	0.18%
D	Count	1,034	672	475	11	2,192
	Row %	47.17%	30.66%	21.67%	0.50%	100.00%
	Column %	77.80%	5.41%	7.25%	4.37%	10.66%
IN	Count	206	10,626	2,045	99	12,976
	Row %	1.59%	81.89%	15.76%	0.76%	100.00%
	Column %	15.50%	85.51%	31.23%	39.29%	63.13%
M	Count	80	703	3,846	55	4,684
	Row %	1.71%	15.01%	82.11%	1.17%	100.00%
	Column %	6.02%	5.66%	58.74%	21.83%	22.79%
O	Count	6	410	169	82	667
	Row %	0.90%	61.47%	25.34%	12.29%	100.00%
	Column %	0.45%	3.30%	2.58%	32.54%	3.24%
Total	Count	1,329	12,427	6,548	252	20,556
	Row %	6.47%	60.45%	31.85%	1.23%	100.00%
	Column %	100.00%	100.00%	100.00%	100.00%	100.00%

Actual

Predicted

SVM Model		D	IN	M	O	Total
D	Count	1,029	114	59	3	1,205
	Row %	85.39%	9.46%	4.90%	0.25%	100.00%
	Column %	77.43%	0.92%	0.90%	1.19%	5.86%
IN	Count	205	11,221	1,563	131	13,120
	Row %	1.56%	85.53%	11.91%	1.00%	100.00%
	Column %	15.43%	90.30%	23.87%	51.98%	63.83%
M	Count	95	1,079	4,922	83	6,179
	Row %	1.54%	17.46%	79.66%	1.34%	100.00%
	Column %	7.15%	8.68%	75.17%	32.94%	30.06%
O	Count	0	13	4	35	52
	Row %	0.00%	25.00%	7.69%	67.31%	100.00%
	Column %	0.00%	0.10%	0.06%	13.89%	0.25%
Total	Count	1,329	12,427	6,548	252	20,556
	Row %	6.47%	60.45%	31.85%	1.23%	100.00%
	Column %	100.00%	100.00%	100.00%	100.00%	100.00%

Actual

Predicted

Ensemble Model		D	IN	M	O	Total
D	Count	1,024	231	189	7	1,451
	Row %	70.57%	15.92%	13.03%	0.48%	100.00%
	Column %	77.05%	1.86%	2.89%	2.78%	7.06%
IN	Count	219	11,503	2,159	152	14,033
	Row %	1.56%	81.97%	15.39%	1.08%	100.00%
	Column %	16.48%	92.56%	32.97%	60.32%	68.27%
M	Count	81	672	4,174	67	4,994
	Row %	1.62%	13.46%	83.58%	1.34%	100.00%
	Column %	6.09%	5.41%	63.74%	26.59%	24.29%
O	Count	5	21	26	26	78
	Row %	6.41%	26.92%	33.33%	33.33%	100.00%
	Column %	0.38%	0.17%	0.40%	10.32%	0.38%
Total	Count	1,329	12,427	6,548	252	20,556
	Row %	6.47%	60.45%	31.85%	1.23%	100.00%
	Column %	100.00%	100.00%	100.00%	100.00%	100.00%

Select Model

			Actual				
	C5 Model		D	IN	M	O	Total
Predicted	D	Count	685	77	38	1	801
		Row %	85.52%	9.61%	4.74%	0.12%	100.00%
		Column %	51.54%	0.62%	0.58%	0.40%	3.90%
	IN	Count	575	11,555	3,004	168	15,302
		Row %	3.76%	75.51%	19.63%	1.10%	100.00%
		Column %	43.27%	92.98%	45.88%	66.67%	74.44%
	M	Count	69	786	3,505	72	4,432
		Row %	1.56%	17.73%	79.08%	1.62%	100.00%
		Column %	5.19%	6.32%	53.53%	28.57%	21.56%
	O	Count	0	9	1	11	21
		Row %	0.00%	42.86%	4.76%	52.38%	100.00%
		Column %	0.00%	0.07%	0.02%	4.37%	0.10%
	Total	Count	1,329	12,427	6,548	252	20,556
		Row %	6.47%	60.45%	31.85%	1.23%	100.00%
		Column %	100.00%	100.00%	100.00%	100.00%	100.00%

			Actual				
	CHAID Model		D	IN	M	O	Total
Predicted	D	Count	519	24	45	1	589
		Row %	88.12%	4.07%	7.64%	0.17%	100.00%
		Column %	39.05%	0.19%	0.69%	0.40%	2.87%
	IN	Count	772	11,940	4,833	176	17,721
		Row %	4.36%	67.38%	27.27%	0.99%	100.00%
		Column %	58.09%	96.08%	73.81%	69.84%	86.21%
	M	Count	38	463	1,670	75	2,246
		Row %	1.69%	20.61%	74.35%	3.34%	100.00%
		Column %	2.86%	3.73%	25.50%	29.76%	10.93%
	Total	Count	1,329	12,427	6,548	252	20,556
		Row %	6.47%	60.45%	31.85%	1.23%	100.00%
		Column %	100.00%	100.00%	100.00%	100.00%	100.00%

			Actual				
	Logistic Regression		D	IN	M	O	Total
Predicted	D	Count	704	255	140	6	1,105
		Row %	63.71%	23.08%	12.67%	0.54%	100.00%
		Column %	52.97%	2.05%	2.14%	2.38%	5.38%
	IN	Count	519	10,261	2,377	123	13,280
		Row %	3.91%	77.27%	17.90%	0.93%	100.00%
		Column %	39.05%	82.57%	36.30%	48.81%	64.60%
	M	Count	95	1,582	3,822	43	5,542
		Row %	1.71%	28.55%	68.96%	0.78%	100.00%
		Column %	7.15%	12.73%	58.37%	17.06%	26.96%
	O	Count	11	329	209	80	629
		Row %	1.75%	52.31%	33.23%	12.72%	100.00%
		Column %	0.83%	2.65%	3.19%	31.75%	3.06%
	Total	Count	1,329	12,427	6,548	252	20,556
		Row %	6.47%	60.45%	31.85%	1.23%	100.00%
		Column %	100.00%	100.00%	100.00%	100.00%	100.00%

			Actual				
	SVM Model		D	IN	M	O	Total
Predicted	D	Count	713	84	44	2	843
		Row %	84.58%	9.96%	5.22%	0.24%	100.00%
		Column %	53.65%	0.68%	0.67%	0.79%	4.10%
	IN	Count	539	11,389	2,792	154	14,874
		Row %	3.62%	76.57%	18.77%	1.04%	100.00%
		Column %	40.56%	91.65%	42.64%	61.11%	72.36%
	M	Count	77	941	3,708	75	4,801
		Row %	1.60%	19.60%	77.23%	1.56%	100.00%
		Column %	5.79%	7.57%	56.63%	29.76%	23.36%
	O	Count	0	13	4	21	38
		Row %	0.00%	34.21%	10.53%	55.26%	100.00%
		Column %	0.00%	0.10%	0.06%	8.33%	0.18%
	Total	Count	1,329	12,427	6,548	252	20,556
		Row %	6.47%	60.45%	31.85%	1.23%	100.00%
		Column %	100.00%	100.00%	100.00%	100.00%	100.00%

			Actual				
	Ensemble Model		D	IN	M	O	Total
Predicted	D	Count	690	75	66	1	832
		Row %	82.93%	9.01%	7.93%	0.12%	100.00%
		Column %	51.92%	0.60%	1.01%	0.40%	4.05%
	IN	Count	572	11,734	3,245	169	15,720
		Row %	3.64%	74.64%	20.64%	1.08%	100.00%
		Column %	43.04%	94.42%	49.56%	67.06%	76.47%
	M	Count	65	612	3,234	74	3,985
		Row %	1.63%	15.36%	81.15%	1.86%	100.00%
		Column %	4.89%	4.92%	49.39%	29.37%	19.39%
	O	Count	2	6	3	8	19
		Row %	10.53%	31.58%	15.79%	42.11%	100.00%
		Column %	0.15%	0.05%	0.05%	3.17%	0.09%
	Total	Count	1,329	12,427	6,548	252	20,556
		Row %	6.47%	60.45%	31.85%	1.23%	100.00%
		Column %	100.00%	100.00%	100.00%	100.00%	100.00%

Cardio Model

Actual

Predicted

C5 Model		D	IN	M	O	Total
D	Count	1,042	73	54	0	1,169
	Row %	89.14%	6.24%	4.62%	0.00%	100.00%
	Column %	78.05%	0.99%	0.89%	0.00%	7.86%
IN	Count	207	6,274	1,404	23	7,908
	Row %	2.62%	79.34%	17.75%	0.29%	100.00%
	Column %	15.51%	84.69%	23.26%	23.71%	53.16%
M	Count	85	1,055	4,566	45	5,751
	Row %	1.48%	18.34%	79.39%	0.78%	100.00%
	Column %	6.37%	14.24%	75.63%	46.39%	38.66%
O	Count	1	6	13	29	49
	Row %	2.04%	12.24%	26.53%	59.18%	100.00%
	Column %	0.07%	0.08%	0.22%	29.90%	0.33%
Total	Count	1,335	7,408	6,037	97	14,877
	Row %	8.97%	49.79%	40.58%	0.65%	100.00%
	Column %	100.00%	100.00%	100.00%	100.00%	100.00%

Actual

Predicted

CHAID Model		D	IN	M	O	Total
D	Count	702	127	83	0	912
	Row %	76.97%	13.93%	9.10%	0.00%	100.00%
	Column %	52.58%	1.71%	1.37%	0.00%	6.13%
IN	Count	415	6,304	1,774	19	8,512
	Row %	4.88%	74.06%	20.84%	0.22%	100.00%
	Column %	31.09%	85.10%	29.39%	19.59%	57.22%
M	Count	218	977	4,180	78	5,453
	Row %	4.00%	17.92%	76.66%	1.43%	100.00%
	Column %	16.33%	13.19%	69.24%	80.41%	36.65%
Total	Count	1,335	7,408	6,037	97	14,877
	Row %	8.97%	49.79%	40.58%	0.65%	100.00%
	Column %	100.00%	100.00%	100.00%	100.00%	100.00%

Actual

Predicted

SVM Model		D	IN	M	O	Total
D	Count	1,055	38	35	0	1,128
	Row %	93.53%	3.37%	3.10%	0.00%	100.00%
	Column %	79.03%	0.51%	0.58%	0.00%	7.58%
IN	Count	200	6,388	1,237	27	7,852
	Row %	2.55%	81.36%	15.75%	0.34%	100.00%
	Column %	14.98%	86.23%	20.49%	27.84%	52.78%
M	Count	79	975	4,751	38	5,843
	Row %	1.35%	16.69%	81.31%	0.65%	100.00%
	Column %	5.92%	13.16%	78.70%	39.18%	39.28%
O	Count	1	7	14	32	54
	Row %	1.85%	12.96%	25.93%	59.26%	100.00%
	Column %	0.07%	0.09%	0.23%	32.99%	0.36%
Total	Count	1,335	7,408	6,037	97	14,877
	Row %	8.97%	49.79%	40.58%	0.65%	100.00%
	Column %	100.00%	100.00%	100.00%	100.00%	100.00%

Actual

Predicted

Ensemble Model		D	IN	M	O	Total
D	Count	1,018	29	34	0	1,081
	Row %	94.17%	2.68%	3.15%	0.00%	100.00%
	Column %	76.25%	0.39%	0.56%	0.00%	7.27%
IN	Count	226	6,531	1,349	20	8,126
	Row %	2.78%	80.37%	16.60%	0.25%	100.00%
	Column %	16.93%	88.16%	22.35%	20.62%	54.62%
M	Count	90	846	4,644	48	5,628
	Row %	1.60%	15.03%	82.52%	0.85%	100.00%
	Column %	6.74%	11.42%	76.93%	49.48%	37.83%
O	Count	1	2	10	29	42
	Row %	2.38%	4.76%	23.81%	69.05%	100.00%
	Column %	0.07%	0.03%	0.17%	29.90%	0.28%
Total	Count	1,335	7,408	6,037	97	14,877
	Row %	8.97%	49.79%	40.58%	0.65%	100.00%
	Column %	100.00%	100.00%	100.00%	100.00%	100.00%

Orthopedic Model

Actual

Predicted

C5 Model		D	IN	M	O	Total
D	Count	0	1	0	0	1
	Row %	0.00%	100.00%	0.00%	0.00%	100.00%
	Column %	0.00%	0.02%	0.00%	0.00%	0.02%
IN	Count	11	5,000	259	109	5,379
	Row %	0.20%	92.95%	4.82%	2.03%	100.00%
	Column %	91.67%	97.87%	56.67%	76.22%	94.02%
M	Count	1	95	197	9	302
	Row %	0.33%	31.46%	65.23%	2.98%	100.00%
	Column %	8.33%	1.86%	43.11%	6.29%	5.28%
O	Count	0	13	1	25	39
	Row %	0.00%	33.33%	2.56%	64.10%	100.00%
	Column %	0.00%	0.25%	0.22%	17.48%	0.68%
Total	Count	12	5,109	457	143	5,721
	Row %	0.21%	89.30%	7.99%	2.50%	100.00%
	Column %	100.00%	100.00%	100.00%	100.00%	100.00%

Actual

Predicted

CHAID Model		D	IN	M	O	Total
D	Count	0	0	0	0	0
	Row %	0.00%	0.00%	0.00%	0.00%	0.00%
	Column %	0.00%	0.00%	0.00%	0.00%	0.00%
IN	Count	12	5,051	326	133	5,522
	Row %	0.22%	91.47%	5.90%	2.41%	100.00%
	Column %	100.00%	98.86%	71.33%	93.01%	96.52%
M	Count	0	58	131	10	199
	Row %	0.00%	29.15%	65.83%	5.03%	100.00%
	Column %	0.00%	1.14%	28.67%	6.99%	3.48%
Total	Count	12	5,109	457	143	5,721
	Row %	0.21%	89.30%	7.99%	2.50%	100.00%
	Column %	100.00%	100.00%	100.00%	100.00%	100.00%

Actual

Predicted

SVM Model		D	IN	M	O	Total
D	Count	0	0	0	0	0
	Row %	0.00%	0.00%	0.00%	0.00%	0.00%
	Column %	0.00%	0.00%	0.00%	0.00%	0.00%
IN	Count	11	4,998	204	92	5,305
	Row %	0.21%	94.21%	3.85%	1.73%	100.00%
	Column %	91.67%	97.83%	44.64%	64.34%	92.73%
M	Count	1	77	248	12	338
	Row %	0.30%	22.78%	73.37%	3.55%	100.00%
	Column %	8.33%	1.51%	54.27%	8.39%	5.91%
O	Count	0	34	5	39	78
	Row %	0.00%	43.59%	6.41%	50.00%	100.00%
	Column %	0.00%	0.67%	1.09%	27.27%	1.36%
Total	Count	12	5,109	457	143	5,721
	Row %	0.21%	89.30%	7.99%	2.50%	100.00%
	Column %	100.00%	100.00%	100.00%	100.00%	100.00%

Actual

Predicted

Ensemble Model		D	IN	M	O	Total
D	Count	0	0	0	0	0
	Row %	0.00%	0.00%	0.00%	0.00%	0.00%
	Column %	0.00%	0.00%	0.00%	0.00%	0.00%
IN	Count	12	5,063	270	114	5,459
	Row %	0.22%	92.75%	4.95%	2.09%	100.00%
	Column %	100.00%	99.10%	59.08%	79.72%	95.42%
M	Count	0	45	187	8	240
	Row %	0.00%	18.75%	77.92%	3.33%	100.00%
	Column %	0.00%	0.88%	40.92%	5.59%	4.20%
O	Count	0	1	0	21	22
	Row %	0.00%	4.55%	0.00%	95.45%	100.00%
	Column %	0.00%	0.02%	0.00%	14.69%	0.38%
Total	Count	12	5,109	457	143	5,721
	Row %	0.21%	89.30%	7.99%	2.50%	100.00%
	Column %	100.00%	100.00%	100.00%	100.00%	100.00%

Appendix 5 - Modeling Nodes Description

SVM – *Machine (SVM) is a robust classification and regression technique that maximizes the predictive accuracy o“Support Vector f a model without overfitting the training data. SVM is particularly suited to analyzing data with very large numbers (for example, thousands) of predictor fields.”* (IBM SPSS Modeler Help, 2015)

CHAID – *“CHAID, or Chi-squared Automatic Interaction Detection, is a classification method for building decision trees by using chi-square statistics to identify optimal splits. CHAID first examines the crosstabulations between each of the input fields and the outcome, and tests for significance using a chi-square independence test. If more than one of these relations is statistically significant, CHAID will select the input field that is the most significant (smallest p value). If an input has more than two categories, these are compared, and categories that show no differences in the outcome are collapsed together. This is done by successively joining the pair of categories showing the least significant difference. This category-merging process stops when all remaining categories differ at the specified testing level. For nominal input fields, any categories can be merged; for an ordinal set, only contiguous categories can be merged.* (IBM SPSS Modeler Help, 2015)

C5 – “*This node uses the C5.0 algorithm to build either a decision tree or a rule set. A C5.0 model works by splitting the sample based on the field that provides the maximum information gain. Each subsample defined by the first split is then split again, usually based on a different field, and the process repeats until the subsamples cannot be split any further. Finally, the lowest-level splits are reexamined, and those that do not contribute significantly to the value of the model are removed or pruned.”* (IBM SPSS Modeler Help, 2015)

Logistic Regression – *“Logistic regression, also known as nominal regression, is a statistical technique for classifying records based on values of input fields. It is analogous to linear regression but takes a categorical target field instead of a numeric*

one. Both binomial models (for targets with two discrete categories) and multinomial models (for targets with more than two categories) are supported.
Logistic regression works by building a set of equations that relate the input field values to the probabilities associated with each of the output field categories. Once the model is generated, it can be used to estimate probabilities for new data. For each record, a probability of membership is computed for each possible output category. The target category with the highest probability is assigned as the predicted output value for that record." (IBM SPSS Modeler Help, 2015)

Ensemble – "*The Ensemble node combines two or more model nuggets to obtain more accurate predictions than can be gained from any of the individual models. By combining predictions from multiple models, limitations in individual models may be avoided, resulting in a higher overall accuracy. Models combined in this manner typically perform at least as well as the best of the individual models and often better."* (IBM SPSS Modeler Help, 2015)

Works Cited

Elizabeth D. Liddy, C. L. (1993). A sublanguage approach to natural language processing for an expert system. *Information Processing & Managament, 29*(5), pp. 633-645.

Elizabeth D. Liddy, S. S. (n.d.). Sublanguage Analysis Applied to Trouble Tickets. *Center for Natural Language Processing*, pp. 752-757.

Gary Miner, J. E. (2012). *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications.* Waltham, MA: Academic Press, Elsevier.

IBM SPSS Modeler Help. (2015, 4 27). Retrieved from IBM SPSS: http://127.0.0.1:58543/help/index.jsp?topic=/com.ibm.spss.modeler.help/clementine/clem_intro.htm

Krippendorff, K. (1989). Content Analysis. *International Encyclopedia of Communication, 1*, pp. 403-407.

Mark Devaney, A. R. (2005). PREVENTING FAILURES BY MINING MAINTENANCE LOGS WITH CASE-BASED REASONING. *59th Meeting of the Society for Machinery Failure Prevention Technology.*

Svetlana Symonenko, S. R. (n.d.). Illuminating Trouble Tickets with Sublanguage Theory. Syracuse, NY, USA: Center for Natural Language Processing, School Of Information Studies, Syracuse University.

U.S. Food and Drug Administration. (2014, November 10). *MAUDE - Manufacturer and User Facility Device Experience*. Retrieved from FDA.gov: http://www.accessdata.fda.gov/scripts/cdrh/cfdocs/cfMAUDE/search.CFM

Yiming Yang, J. O. (1997). A Comparative Study on Feature Selection in Text Categorization. *ICML '97 Proceedings of the Fourteenth International Conference on Machine Learning* (pp. 412-420). San Francisco: Morgan Kaufmann Publishers Inc. .