

Drug-Induced Liver Injury (DILI) Classification using US Food and Drug Administration (FDA)-Approved Drug Labeling and FDA Adverse Event Reporting System (FAERS) data

Qais Hatim¹, Lilliam Rosario¹, Eileen Navarro Almario¹, Kendra Worthy¹, Tom Sabo², Emily McRae²

1 U.S. Food and Drug Administration, 2 SAS Institute Inc.

Research Problems

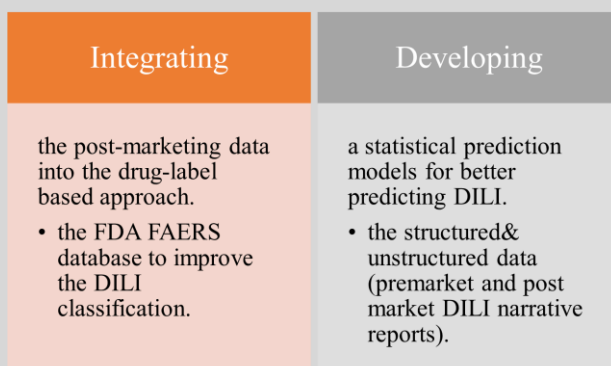
- Why is it valuable to define **DILI**?
 - Defining DILI is challenging as:
 - It is a diagnosis of exclusion, the incidence and severity may be drug specific
- Biomarkers and methodologies are being developed to assess hepatotoxicity but:
 - require a list of drugs with well-annotated DILI potential
- What do we get from assessing **hepatotoxicity**?
 - A drug classification scheme is essential to evaluate the performance of existing DILI biomarkers and discover novel DILI biomarkers but:
 - no adopted practice can classify a drug's DILI potential in humans.
- Do we label properly to save lives?
 - Drug labels used to develop a systematic and objective classification scheme Rule-of-two (RO2). However:
 - highly context specific
 - rarity of DILI in the premarket experience
 - the complex phenotypes of DILI
 - drugs are often used in combination with other medications.

Research Solution

This research aimed to enrich the RO2 model based on **machine learning and data-mining modeling** of premarket and post market DILI narrative reports. We utilized the FDA FAERS database that provides post-marketing surveillance data in order to improve the DILI classification by integrating the post-marketing data into the drug-label based approach to further improve the accuracy of DILI classifications. This research will develop a statistical prediction model for better predicting DILI in humans.

Therefore, by **Applying computational** approaches in the analysis of FAERS, we will advance the understanding of the feasibility of applying these advance techniques to uncover relationships between drugs and hepatic failure. Outputs offer information on:

- the drug combinations related to reported hepatic failure
- clarify factors that can predict a greater degree of the hepatic failure from serious (death) to minor events (treatable)
- rate the serious impact of using these drugs on patients.

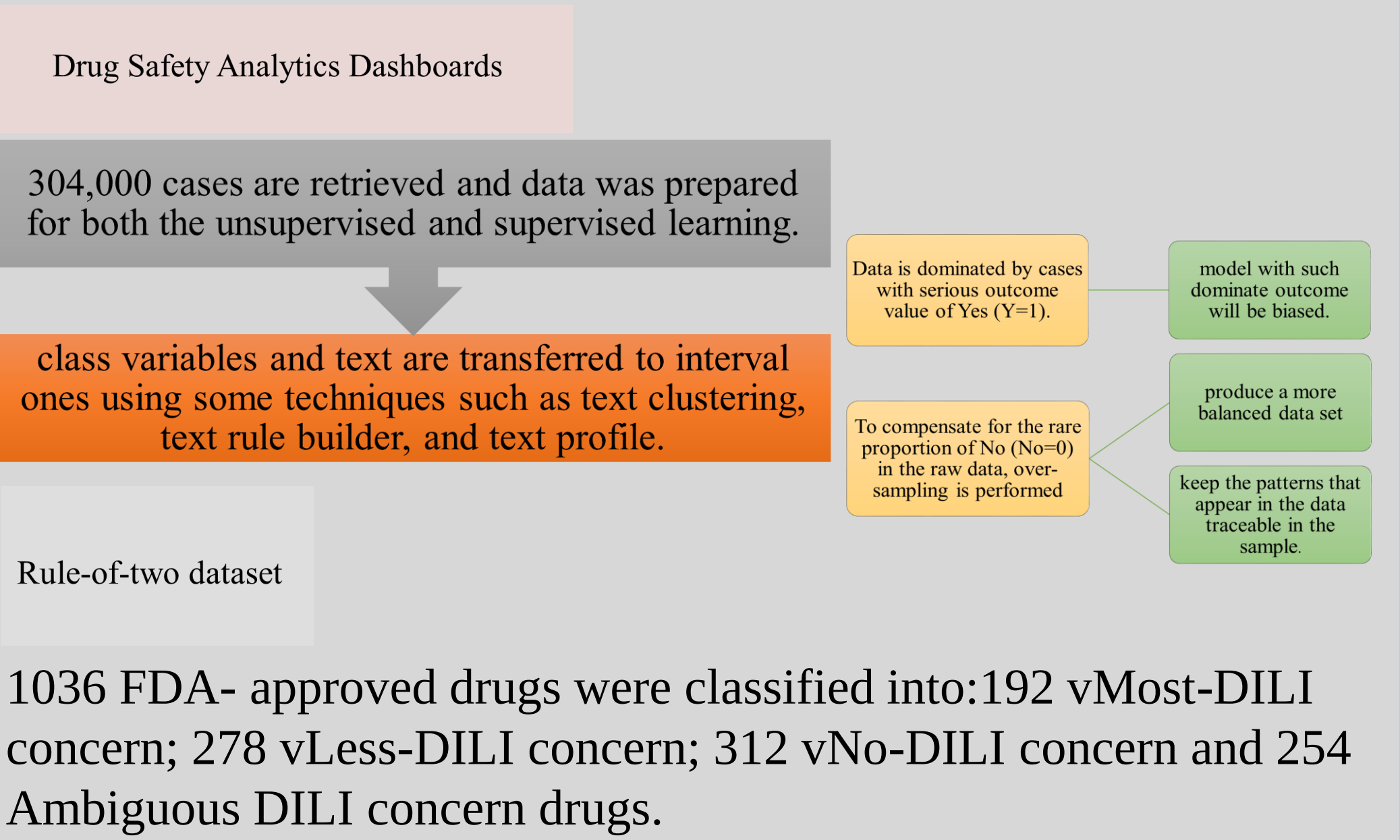


Methodology



Data Extraction & Preprocessing

- Empirica Signal
- Empirica Signal served as the source of data retrieval based on the preferred term (PT) or standard MedDRA query (SMQ) equal to 'Drug related hepatic disorders - severe events only (SMQ) [narrow]'. Prioritizing investigations might be based on scores for statistical significance, rather than for association, to avoid following up potential associations that could have arisen merely by chance.
 - However, unnecessary focus on drugs and events that are common overall in the database can be an outcome of using a PRR or ROR p-value to rank associations.
 - Therefore, **in this research**, we prioritize investigations based on both **significance and association scores**, rather than relying on only one score. A threshold of EBGM>2 and EB05>1 are used and therefore the number of cases reduced to only 14,436 cases from the initial retrieved 171,890 cases.



1036 FDA- approved drugs were classified into:192 vMost-DILI concern; 278 vLess-DILI concern; 312 vNo-DILI concern and 254 Ambiguous DILI concern drugs.

Model Building & Domain Experts

Analytics Application/Association Analysis

Association analysis is a popular technique that is used to identify and visualize relationships (association) between different objects. Query could be nontrivial to be answered manually with big dataset. For example, what linkage of DILI preferred terms (events) can be observed from post-market data? Association analysis can address such relationship by defining: association rules and calculating the support for the combination of the PTs.

Three scenarios are developed for the subset data from Empirica Signal (i.e., the 14,436 cases after utilizing both significance and association scores in prioritizing DILI cases). For each scenario, association model is built based on different settings for *minimum support*, *minimum confidence*, *minimum lift*, *maximum antecedents*, and *maximum rule size*. Resetting these values allow us to cover more association rules as well as understand the optimal setting that provides more informative rules.

Data Sets Aggregation

This research data has two different domains (i.e., pre-marketing and post-marketing report).

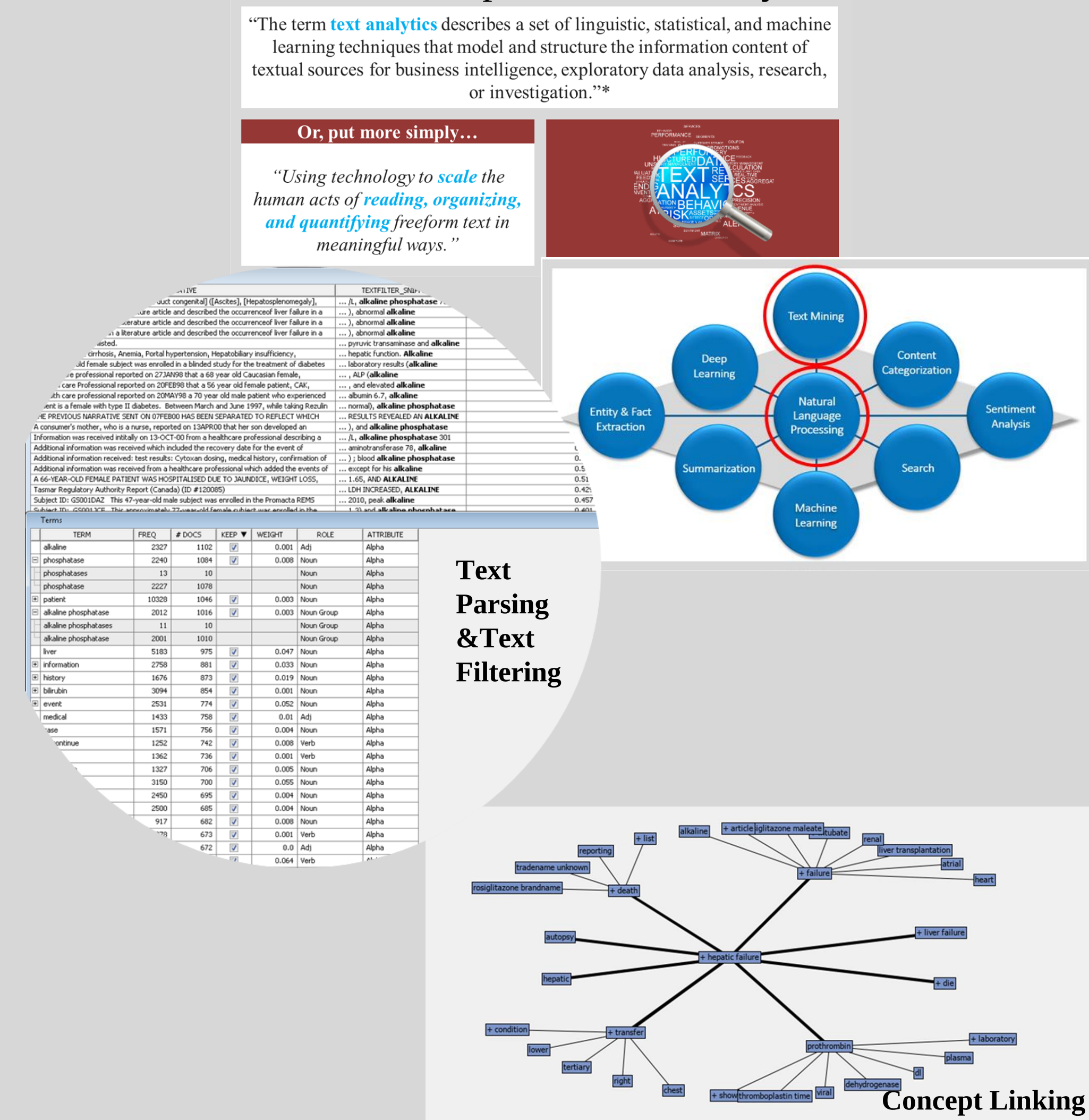
Drug corpus from RO2 classifying DILI risk based on drug labeling

Empirica Signal and Drug Safety Analytics Dashboards are based on FAERS data which is post-marketing data.

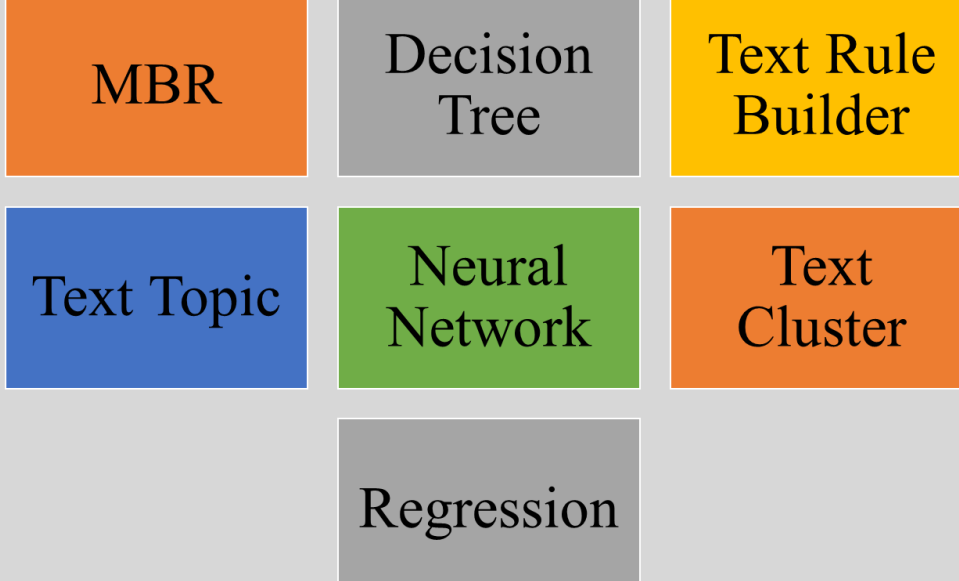
Numerous customized SQL were developed to match the RO2 compound names (1036 unique drugs) with 182474 DILI cases from FAERS.

Text Analytics

To meet the challenges posed by unstructured text, text mining, which employs a wide range of statistical, machine learning, and linguistic techniques for natural language processing (NLP), is utilized in this paper. Moreover, while some of the FAERS information is structured, a significant portion of it remains in narrative formats. Much of the information that is critical to risk assessments, such as signs and symptoms, disease status and severity, and medical history, are typically imbedded in narrative text. Accordingly, we exploited both the structured and unstructured FAERS data items and developed different analytical models.

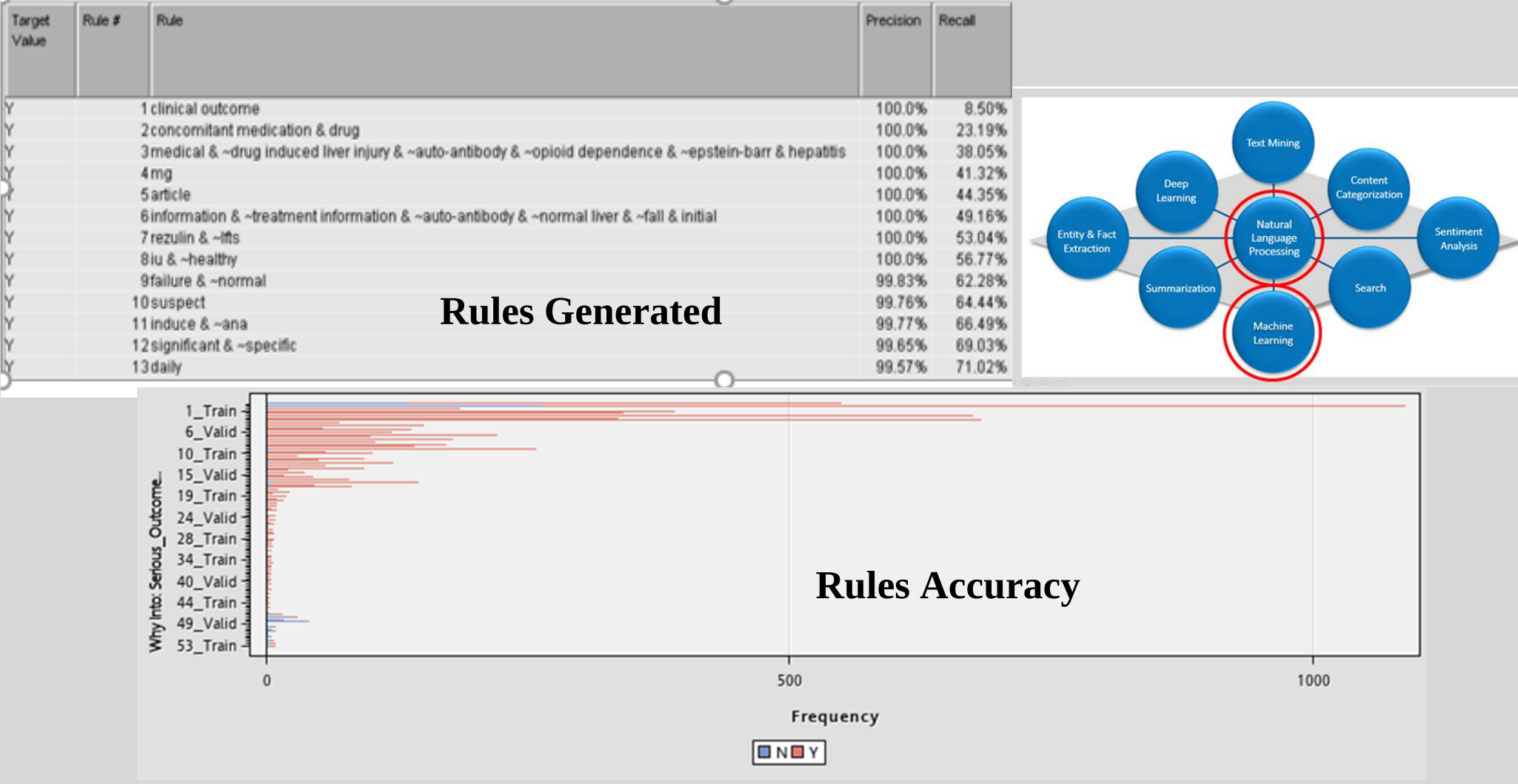


Predictive Analysis/Supervised & Unsupervised Models



Text Categorization-Rule Building

Boolean rules are created from small subsets of terms to predict a categorical target variable, in our case the binary serious outcome. To prevent overtraining, control the complexity of rules, and determine the exhaustiveness of the rule search process, different settings for these specifications were tried and several analyses were compared to set the optimal text rule builder setting.



Text Categorization-Clustering

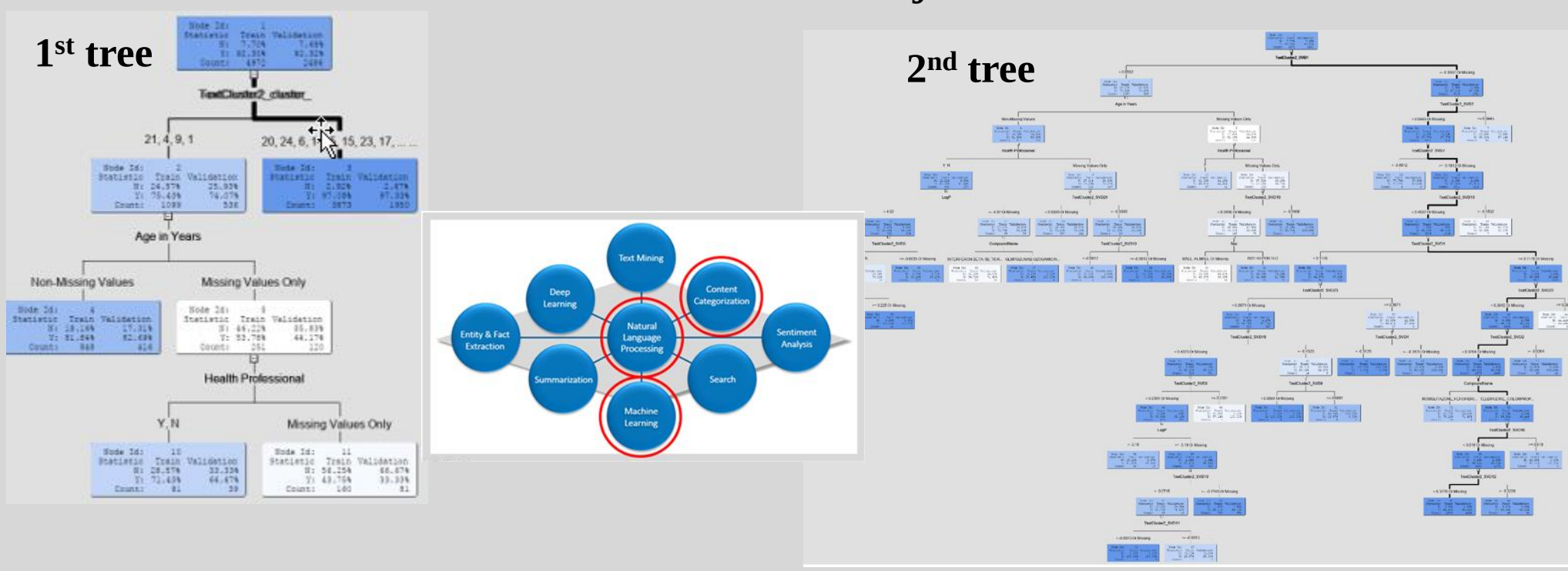
The text cluster uses a Expectation-Maximization clustering algorithm to describe the contents clusters. The primary purpose of the text cluster is to derive clusters. This is achieved by deriving a numeric representation for each document. Latent Semantic Analysis (LSA), implemented through Singular Value Decomposition (SVD), produces the numeric representation and the derived numeric representation produces the clusters. The SVD is applied to approximate the term/document matrix, and documents are projected into a reduced dimensional space. The generated SVD dimensions are those that fit the subspace the best in terms of least-square best fit.



Supervised Model-Decision Tree

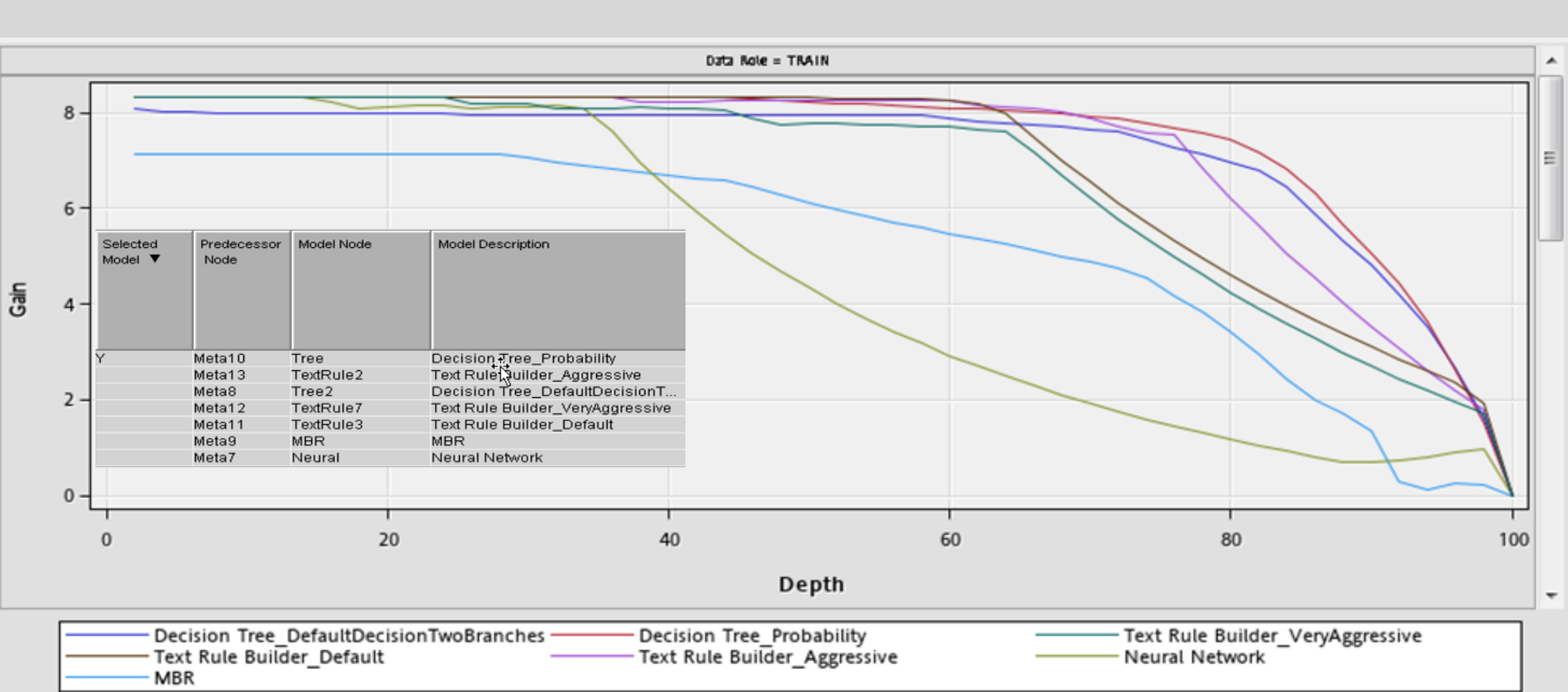
We utilized decision tree to perform the three essential tasks that predictive models perform which are predict new liver failure cases, select useful inputs, and optimize complexity. Decision trees, similar to other modeling methods, address each of the modeling essentials. Cases are scored using prediction rules. A split-search algorithm facilitates input selection. Model complexity is addressed by pruning. The output from the cluster analysis is the input to the decision tree modeling. Two decision tree models have been developed.

1st tree: the SVD numeric values have been rejected only the nominal values of cluster numbers will input the decision tree modeling with other FAERS input variables. **2nd tree:** the SVDs is utilized as input to the decision tree with other FAERS variables and cluster number variable has been rejected.



Model Comparison

The model comparison tool facilitates selection between model tools using the summary statistics. Summary statistics transform model performance to numerical scores. The model comparison tool collates statistics from different modeling nodes for easy comparison. Out of the 8 models that have developed in this research, the winning model is the decision tree which is indicated with letter Y as winning model.



Scoring

To see how accurately the champion model classifies new data that do not contain a target value (i.e., serious outcome), 4000 cases have been selected randomly and assigned the role of score data set. The score table generated that illustrates the model's predicted FAERS serious outcome category and the model probability for this category (probability of classification).