Project: Acquired

A Look Into an Untapped Resource

Brady Coye, Drew Rice, Johnathon Schultz

Introduction

In data analytics, the desired data is not always available. The focus of our project was to obtain information not freely available to the public. Our term project was centered around the overlaying of difficult to acquire traffic incident data points on top readily available, yet separate, weather data. We sought to implement basic machine learning algorithms and data visualization to understand this potentially unexplored field of study for Greenville, SC.

We used data mining to build a model that allowed us to classify local traffic incidents, based on previous, similar traffic incidents and weather patterns. The main challenges were in the data acquisition and preparation stage, in which datasets from multiple sources were scraped and combined. Preprocessing involved several steps, including scraping the data, cleaning dependent attributes, handling missing values, and converting to the correct format. Results were as anticipated, with inclement weather conditions leading to more traffic incidents. Visualizing the data gave further insight into correlation between traffic incidents and weather patterns. However, our underlying results were that new data points, including proprietary or ungathered, may be made available and used to produce meaningful output given the appropriate motivation.

Dataset description

The crime data was considered from http://goo.gl/2ROhdm
The weather data was obtained from http://goo.gl/VTt1kk
The traffic data was gathered from http://goo.gl/BG3RnL

Month	Month the data was recorded on
Day	Day of month the data was recorded on
Hour	Hour of day the data was recorded on
Minute	Minute of the hour data was recorded on
Sky Condition	Description of the appearance of the sky (Clear or Not Clear)
Temperature	Degree of hotness or coldness of the air

Precip	Amount of any form of water falling from the sky in inches
Humidity	Percentage of water vapor concentration in the air
Visibility	Distance an object can be clearly discerned in miles
Wind speed	Speed of wind in MPH
Status	The general categorical state of the incident instance
Incident Type	The classification of the incident instance

Data acquisition

Traffic Incident Data

In order to obtain the traffic incident data, several steps were required. First, an API stream to the live incident webpage was created so that we could programmatically poll at a regular interval to save the stream of information. On the back end, a database was created to hold each traffic incident instance and all of its attributes. Finally, we implemented a task to convert the API JSON data stream into database instances; then, wrote a CRON job to fire the task on a minute by minute basis. The result is a down to the minute log of Greenville traffic incidents.

Weather Data

The weather data was pre-formatted in CSV format. We downloaded weather data broken down hourly for each month used in analysis.

Crime Data

Was unable to be obtained. Though many sources and methods were deployed to attempt to obtain this information, crime information for Greenville is protected and not easily scrapable. Proprietary crime data was available but at a high cost and not feasible for the scope of this project.

Data preparation

Traffic Incident Data

In an effort to remove redundant or non predictive data, several attributes were modified or removed altogether. The unique key value was removed to avoid overfitting. We separated the timestamp into year, month, day, hour and minute attributes by removing delimiters, time zone information, year (since it was always 2016), and seconds (since they were always 00). EndTime was removed altogether because the attribute was never populated and the County attribute was removed because all incidents occurred in Greenville county.

Weather Data

For each month, a lot of modifications were done to remove redundant, missing, or non predictive data. All of the metadata about the station prefacing the weather instances were removed. Attributes with the same or empty values across all instances were removed entirely such as: WBAN, StationType, all attributes denoted as a “Flag”, PressureTendancy, PressureChange, SeaLevelPressure, RecordType, and WeatherType. We separated the timestamp into year, month, day and removed the year attribute (since it was always 2016). Then we split the time into hour and minute in Excel:

Hour	=VALUE(IF(LEFT(C2,LEN(C2)-2)="",0,LEFT(C2,LEN(C2)-2)))
Minute	=VALUE(RIGHT(C2, 2))

Compilation

At this point, we had the data formatted with several time attributes in common. Both data sets were loaded into Tableau to left join the weather instances to the traffic instances. We then exported the resulting CSV from Tableau and manually converted the file to ARFF. We did this separately for March and April to produce two separate files so that we had March available as training data and were able to load April instances in as test data. Most attributes are numeric, however, for SkyCondition, Status, and IncidentType we had to extrapolate all possible values for the ARFF attribute heading since the attributes are nominal.

Data analysis

Our hypothesis was that the weather can predict the type of traffic accidents. We initially used unsupervised association learning in order to test our hypothesis. The Apriori algorithm was used to give us rules as evidence to support our hypothesis. The results from the visualization are fairly intuitive but the change in the exact number of incidents was interesting to discover. Though the preprocessing was extremely tedious and time consuming, the actual data analysis was a process of stumbling across interesting relationships between attributes.

Prior to running machine learning algorithms in WEKA, we sought to understand the dataset further with data visualization in Tableau; Tableau was the main source of our visual analysis. The application is an intuitive, powerful tool that allows users to import a dataset and discover a story that wouldn’t be understood without visualization. Visualization techniques are imperative to gaining insight into any dataset, especially in terms of big data. Understanding trends or correlations is nearly impossible when looking at the raw data. When visualization is applied, the underlying relationships between attributes becomes much more clear.

In the case of the weather and traffic incident data, Tableau made understanding the hidden knowledge very easy. Simply by selecting certain attributes and applying different data visualization methods, we were able to determine much about which weather attributes were highly correlated with fluctuations in the number of incidents on a given day.

Results

To find the association rules, the Apriori algorithm was used. The Apriori algorithm iteratively reduces the minimum support until it finds the required number of rules with the given minimum confidence, or a measure of the algorithms accuracy. In our case, the data without a third dataset produced intuitive rules. However, the wonderful weather in Greenville during the months of March and April created a discrepancy in our analysis. The high frequency of days with great weather resulted in overfitting, meaning the model describes random error or noise instead of the underlying relationship; the algorithms used found an overall high number of traffic incidents occurring on days with decent driving conditions. Obviously, the results are overfit to the dataset because, as the upcoming visualizations suggest, more incidents occur when the weather is not ideal for driving. Observations were difficult to make in this circumstance because Greenville has had great weather in both March and April. As a result, most of the accidents occurred on days with no precipitation, high wind, or low visibility. If Greenville had more bad weather, the rules given by the Apriori algorithm would be much more interesting. The most relevant finding from Apriori was:

```
Incident Type=COLLISION:NO INJURY 2157 ==> Precip='(-inf-0.005]' 2130    conf:(0.99)
```

This rule suggests that collisions were less likely to have injuries when it does not rain.

Apriori returned insightful but, in many cases, obviously overfit results. We then turned to classification algorithms to predict the type of accident based solely on the weather. First, ZeroR and OneR were run. ZeroR attempts to classify the instances based on the percentage split of the class attribute, which resulted in a 42.99% accuracy. OneR looks for the most predictive single attribute in the dataset; we found that OneR had 45.81%. These low accuracy percentages meant that further classification was needed.

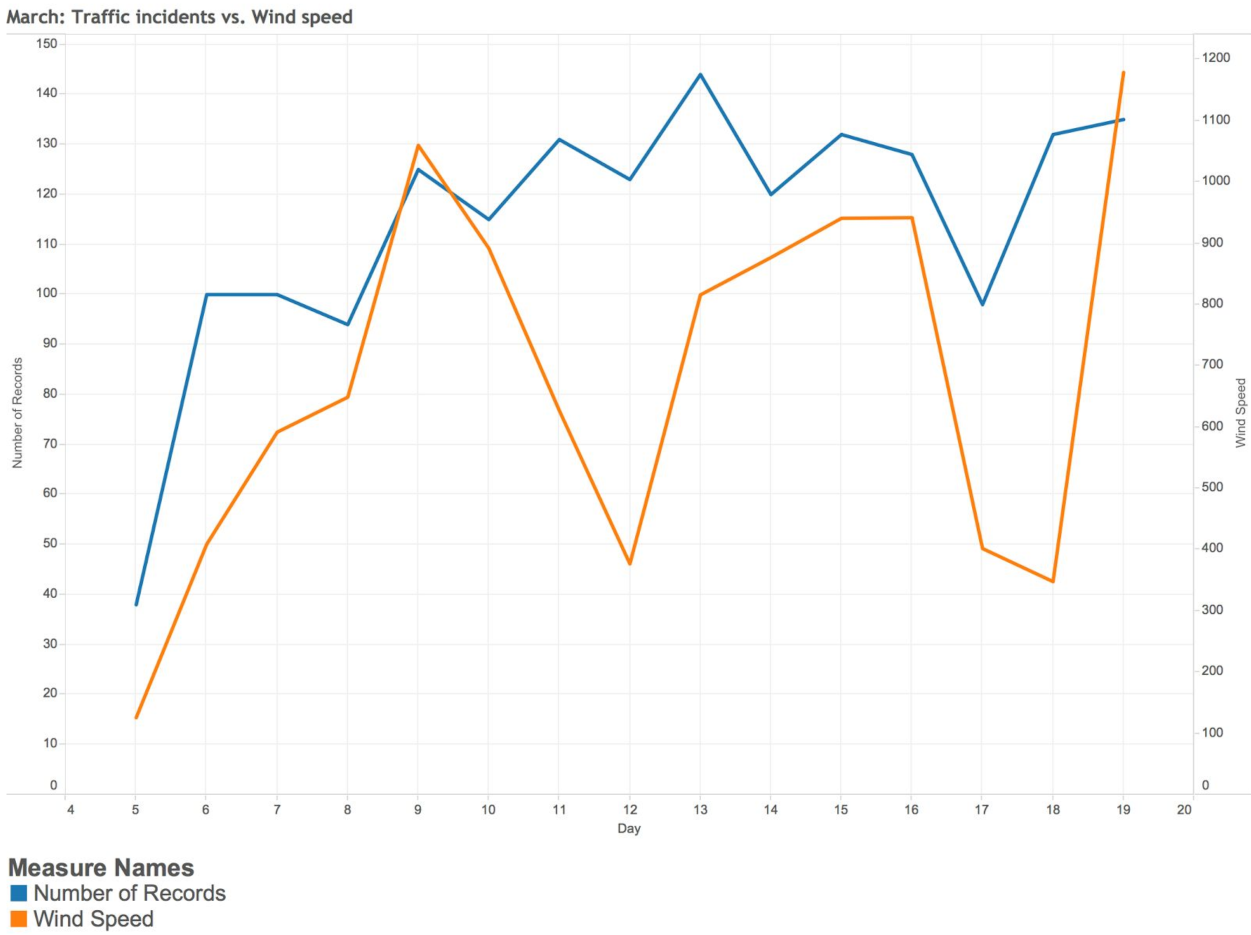
J48, a decision-tree building algorithm, was used next. This algorithm resulted in a very complex tree with the size of the tree being 858 nodes. This immense decision tree resulted in an accuracy of 57.83%. It was clear that further investigation of classification learning was necessary.

Next, we ran nearest neighbor with a *k*, or neighborhood, of 3. This algorithm seeks to classify instances based on their Euclidean distance from other, similar instances. This resulted in an accuracy of 58.33%.

Finally, ensemble learning was the clear answer to improve our model's accuracy. We combined the three previously mentioned classification algorithms and added RandomTree. These four algorithms running in unison give us our best accuracy of all. The ensemble learning algorithm takes the results from each sub-algorithm and averages the probabilities of being correct. This resulted in an overall accuracy of 64.79%.

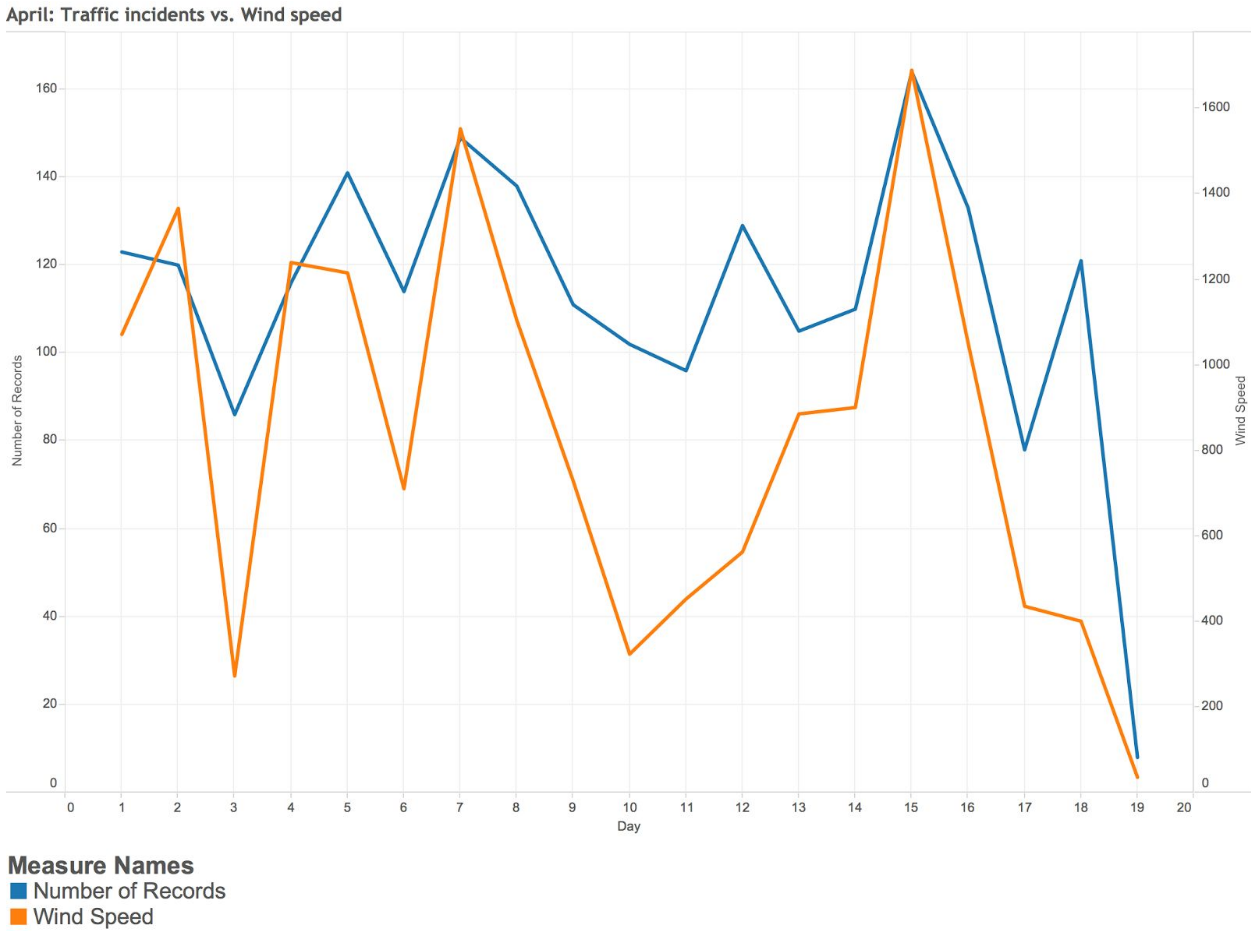
Though these accuracy percentages seem somewhat low at first glance, it is clear that they are a step in the right direction for traffic incident prediction. The model attempts to classify 26 attributes, which is a huge burden for any algorithm. Correctly predicting 64.79% of instances based solely on the weather patterns is remarkable. To reiterate, simply guessing with 26 attributes would give an accuracy of 3.85%. Ensemble learning is clearly the best method to predict traffic incidents. However, take this analysis with a grain of salt; the results are hopelessly optimistic. Using split training and test data gives an accuracy of 29.29% in the best case scenario.

Figure 1:



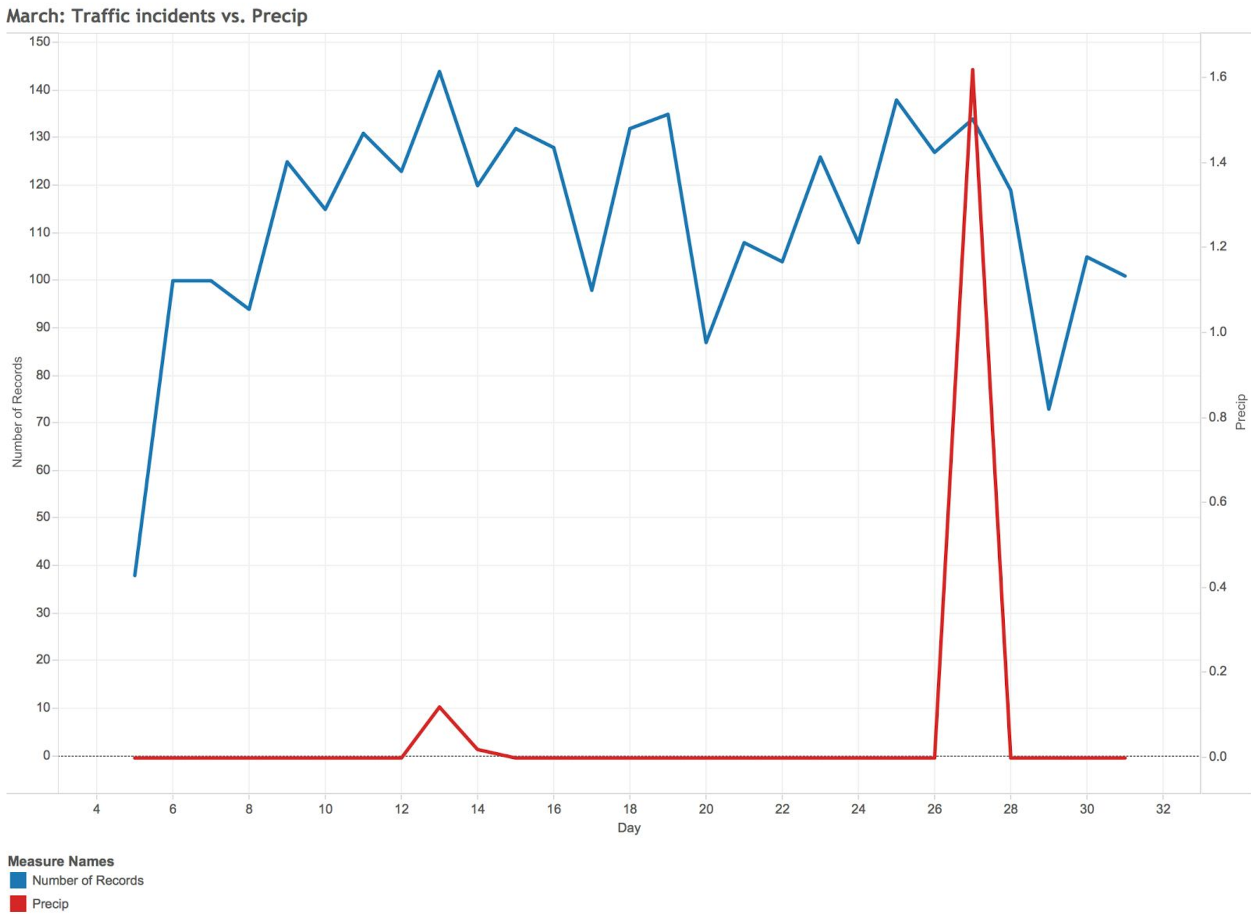
High wind speeds can be characteristic of bad weather and, therefore, bad driving. Though that was simply a hypothesis beforehand, wind speed and traffic incidents are highly correlated. Increases in wind speed clearly relate to an increased number of incidents.

Figure 2:



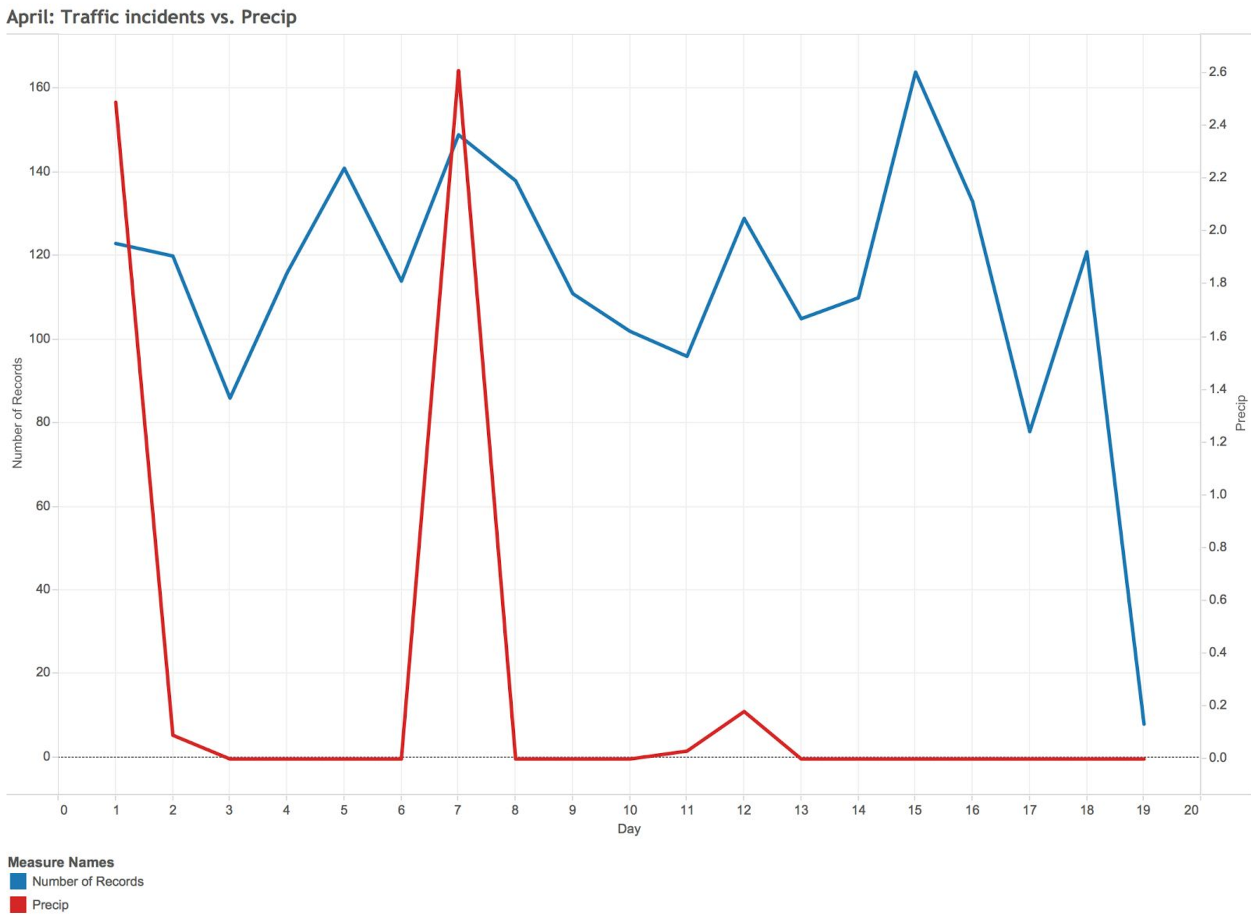
The same relationship seen in the March visualization holds true in April. It seems April has more drastic changes in wind speed throughout the measured time. The number of traffic incidents is affected by the increase in wind speed. Conversely, a low wind speed generally means a decrease in incidents or accidents. In many cases, a high wind speed is a sign of weather not ideal for driving. Therefore, the increase in accidents due to wind speed is simply a strong correlation, not causation. When analyzing a symphony of different attributes, locating causation is generally not the goal of the analysis. Though it is possible that causation can be learned from a dataset, correlations and relationships between attributes are generally the output of the machine learning algorithms.

Figure 3:



In March, there is a clear relationship between when precipitation occurs and number of accidents. Even a small rain, such as on March 13th, causes an increase in traffic incidents. Based on this notion, the same trend should exist in April.

Figure 4:



Precipitation occurring on the 7th and 12th causes dramatic increases in the number of incidents. Again, even a small rain causes drivers' performances to decline. The correlation between precipitation and number of accidents is strong, as the visualization suggests.

Conclusions

The focal point of this research is to combine existing datasets for a study of Greenville traffic and weather patterns. The fusing of two scraped datasets was the process by which we created a unique weather pattern analysis. This method of analysis could be extrapolated to the nation's traffic incidents and weather patterns. In other words, scaling this research to a state or national level would only increase awareness and driving safety.

Experimenting with a large number of algorithms was essential to gathering the best results. Though these results were not necessarily as accurate as we had hoped, that fact is a finding in itself. Our model saw associations that were overfit, where visualization gave a much better insight into the dataset. Visualization is a much quicker and more efficient way to access the hidden knowledge of a dataset, and this can be seen by our March vs. April visualizations.

The preprocessing portion of this project was substantial, allowing us to experiment with combining data from multiple sources. The takeaway from this research emphasizes that previously separate and unavailable datasets can be fused together to provide a greater insight; the combined data provides analysis is greater than the sum of its parts.

Appendix

API polling task written in GO:

```
package main

import (
	"encoding/json"
	// "fmt"
	"io"
	"net/http"
	"time"

	"appengine"
	"appengine/urlfetch"
	"appengine/datastore"
)

const incidentKind = "Incidents"

type APIResponse struct {
	Results Incidents `json:"results"`
}

type Incidents []Incident

type Incident struct {
	IncidentType string `json:"incident_type"`
	StartTime string `json:"start_time"`
	EndTime string `json:"end_time"`
	County string `json:"county"`
	Location string `json:"location"`
	Status string `json:"status"`
}

func defaultIncidentList(c appengine.Context) *datastore.Key {
	return datastore.NewKey(c, incidentKind, "default", 0, nil)
}
```

API polling task written in GO (continued):

```
func timeStringToLocalTime (sourceTime string) string {
	location, _ := time.LoadLocation("EST")
	now := time.Now().In(location)
	formattedTime, _ := time.ParseInLocation("15:04", sourceTime, location)
	formattedTime = time.Date(
			now.Year(),
			now.Month(),
			now.Day(),
			formattedTime.Hour(),
			formattedTime.Minute(),
			formattedTime.Second(),
			formattedTime.Nanosecond(),
			location,
	)
	return formattedTime.Format(time.RFC3339)
}

func (p *Incident) key(c appengine.Context) *datastore.Key {
	p.StartTime = timeStringToLocalTime(p.StartTime)
	return datastore.NewKey(c, incidentKind, p.StartTime+"; "+p.Location, 0, defaultIncidentList(c))
}

func (p Incidents) save(c appengine.Context) (Incidents, error) {
	for i := 0; i < len(p); i++ {
			_, err := datastore.Put(c, p[i].key(c), &p[i])
			if err != nil {
					return nil, err
			}
	}

	return p, nil
}

func decodeIncidents(r io.ReadCloser) (Incidents, error) {
	defer r.Close()
	var apiResponse APIResponse
	err := json.NewDecoder(r).Decode(&apiResponse)
	incidents := apiResponse.Results
	return incidents, err
}

func requestIncidentsFromAPI(url string, ctx appengine.Context) (Incidents, error) {
	client := urlfetch.Client(ctx)
	r, err := client.Get(url)
	if err != nil {
			return nil, err
	}
	incidents, err := decodeIncidents(r.Body)
	return incidents, err
}
```

API polling task written in GO (continued):

```
func handleIncidents(c appengine.Context, r *http.Request) (interface{}, error){
	url :=
"http://api.import.io/store/connector/329b68c2-d010-4ba1-997b-9139806ccaeb/_query?input=webpage/url:http%3A%2F
%2Fafc5102.scdps.gov%2FSCDPS_Exweb%2FSCDPS%2FHighwayPatrol%2FWebCad_External%3FId%3D3&&_api
key=303434a0d9cf4c70b62e016a0736ad949f333a7bdf4c74ec837aae11bbbe7c7086cb08fafe48ff48b46f5b5237603dd
174b1c5f54e82ef4209cd7605aa3068b64a14fb039c1ccdfd0136e97f5e056db5"
	incidents, err := requestIncidentsFromAPI(url, c)
	if err != nil {
		return nil, err
	}
	return incidents.save(c)
}

func handler(w http.ResponseWriter, r *http.Request) {
	c := appengine.NewContext(r)

	val, err := handleIncidents(c, r)
	if err == nil {
		err = json.NewEncoder(w).Encode(val)
	}
	if err != nil {
		c.Errorf("Post Error: %#v", err)
		http.Error(w, err.Error(), http.StatusInternalServerError)
		return
	}
}

func init() {
	http.HandleFunc("/", handler)
}
```

Apriori with discretized values:

1. Humidity='(42-71]' 1908 ==> Precip='(-inf-0.005]' 1908   conf:(1)
2. Humidity='(42-71]' Visibility='(6.75-inf)' 1905 ==> Precip='(-inf-0.005]' 1905   conf:(1)
3. Humidity='(-inf-42]' 1865 ==> Precip='(-inf-0.005]' 1865   conf:(1)
4. Humidity='(-inf-42]' 1865 ==> Visibility='(6.75-inf)' 1865   conf:(1)
5. Humidity='(-inf-42]' Visibility='(6.75-inf)' 1865 ==> Precip='(-inf-0.005]' 1865   conf:(1)
6. Precip='(-inf-0.005]' Humidity='(-inf-42]' 1865 ==> Visibility='(6.75-inf)' 1865   conf:(1)
7. Humidity='(-inf-42]' 1865 ==> Precip='(-inf-0.005]' Visibility='(6.75-inf)' 1865   conf:(1)
8. Humidity='(42-71]' Wind Speed='(-inf-7.333333]' 1111 ==> Precip='(-inf-0.005]' 1111   conf:(1)
9. Humidity='(42-71]' Visibility='(6.75-inf)' Wind Speed='(-inf-7.333333]' 1108 ==> Precip='(-inf-0.005]' 1108   conf:(1)
10. Temperature='(71-inf)' 974 ==> Precip='(-inf-0.005]' 974   conf:(1)
11. Temperature='(71-inf)' 974 ==> Visibility='(6.75-inf)' 974   conf:(1)
12. Temperature='(71-inf)' Visibility='(6.75-inf)' 974 ==> Precip='(-inf-0.005]' 974   conf:(1)
13. Temperature='(71-inf)' Precip='(-inf-0.005]' 974 ==> Visibility='(6.75-inf)' 974   conf:(1)
14. Temperature='(71-inf)' 974 ==> Precip='(-inf-0.005]' Visibility='(6.75-inf)' 974   conf:(1)
15. Temperature='(58-71]' Humidity='(-inf-42]' 955 ==> Precip='(-inf-0.005]' 955   conf:(1)
16. Temperature='(58-71]' Humidity='(-inf-42]' 955 ==> Visibility='(6.75-inf)' 955   conf:(1)
17. Temperature='(58-71]' Humidity='(-inf-42]' Visibility='(6.75-inf)' 955 ==> Precip='(-inf-0.005]' 955   conf:(1)
18. Temperature='(58-71]' Precip='(-inf-0.005]' Humidity='(-inf-42]' 955 ==> Visibility='(6.75-inf)' 955   conf:(1)
19. Temperature='(58-71]' Humidity='(-inf-42]' 955 ==> Precip='(-inf-0.005]' Visibility='(6.75-inf)' 955   conf:(1)
20. Humidity='(-inf-42]' Wind Speed='(-inf-7.333333]' 868 ==> Precip='(-inf-0.005]' 868   conf:(1)

Apriori with discretized values (continued):

21. Humidity='(-inf-42]' Wind Speed='(-inf-7.333333]' 868 ==> Visibility='(6.75-inf)' 868 conf:(1)
22. Humidity='(-inf-42]' Visibility='(6.75-inf)' Wind Speed='(-inf-7.333333]' 868 ==> Precip='(-inf-0.005]' 868 conf:(1)
23. Precip='(-inf-0.005]' Humidity='(-inf-42]' Wind Speed='(-inf-7.333333]' 868 ==> Visibility='(6.75-inf)' 868 conf:(1)
24. Humidity='(-inf-42]' Wind Speed='(-inf-7.333333]' 868 ==> Precip='(-inf-0.005]' Visibility='(6.75-inf)' 868 conf:(1)
25. Incident Type=COLLISION:NO INJURY Humidity='(-inf-42]' 827 ==> Precip='(-inf-0.005]' 827 conf:(1)
26. Incident Type=COLLISION:NO INJURY Humidity='(-inf-42]' 827 ==> Visibility='(6.75-inf)' 827 conf:(1)
27. Incident Type=COLLISION:NO INJURY Humidity='(-inf-42]' Visibility='(6.75-inf)' 827 ==> Precip='(-inf-0.005]' 827 conf:(1)
28. Incident Type=COLLISION:NO INJURY Precip='(-inf-0.005]' Humidity='(-inf-42]' 827 ==> Visibility='(6.75-inf)' 827 conf:(1)
29. Incident Type=COLLISION:NO INJURY Humidity='(-inf-42]' 827 ==> Precip='(-inf-0.005]' Visibility='(6.75-inf)' 827 conf:(1)
30. Temperature='(45-58]' Humidity='(42-71]' 797 ==> Precip='(-inf-0.005]' 797 conf:(1)
31. Temperature='(45-58]' Humidity='(42-71]' 797 ==> Visibility='(6.75-inf)' 797 conf:(1)
32. Temperature='(45-58]' Humidity='(42-71]' Visibility='(6.75-inf)' 797 ==> Precip='(-inf-0.005]' 797 conf:(1)
33. Temperature='(45-58]' Precip='(-inf-0.005]' Humidity='(42-71]' 797 ==> Visibility='(6.75-inf)' 797 conf:(1)
34. Temperature='(45-58]' Humidity='(42-71]' 797 ==> Precip='(-inf-0.005]' Visibility='(6.75-inf)' 797 conf:(1)
35. Humidity='(-inf-42]' Wind Speed='(7.333333-14.666667]' 780 ==> Precip='(-inf-0.005]' 780 conf:(1)
36. Humidity='(-inf-42]' Wind Speed='(7.333333-14.666667]' 780 ==> Visibility='(6.75-inf)' 780 conf:(1)
37. Humidity='(-inf-42]' Visibility='(6.75-inf)' Wind Speed='(7.333333-14.666667]' 780 ==> Precip='(-inf-0.005]' 780 conf:(1)
38. Precip='(-inf-0.005]' Humidity='(-inf-42]' Wind Speed='(7.333333-14.666667]' 780 ==> Visibility='(6.75-inf)' 780 conf:(1)
39. Humidity='(-inf-42]' Wind Speed='(7.333333-14.666667]' 780 ==> Precip='(-inf-0.005]' Visibility='(6.75-inf)' 780 conf:(1)
40. Humidity='(42-71]' 1908 ==> Visibility='(6.75-inf)' 1905 conf:(1)
41. Precip='(-inf-0.005]' Humidity='(42-71]' 1908 ==> Visibility='(6.75-inf)' 1905 conf:(1)
42. Humidity='(42-71]' 1908 ==> Precip='(-inf-0.005]' Visibility='(6.75-inf)' 1905 conf:(1)
43. Humidity='(42-71]' Wind Speed='(-inf-7.333333]' 1111 ==> Visibility='(6.75-inf)' 1108 conf:(1)
44. Precip='(-inf-0.005]' Humidity='(42-71]' Wind Speed='(-inf-7.333333]' 1111 ==> Visibility='(6.75-inf)' 1108 conf:(1)
45. Humidity='(42-71]' Wind Speed='(-inf-7.333333]' 1111 ==> Precip='(-inf-0.005]' Visibility='(6.75-inf)' 1108 conf:(1)
46. Incident Type=COLLISION:NO INJURY Visibility='(6.75-inf)' 2088 ==> Precip='(-inf-0.005]' 2070 conf:(0.99)
47. Incident Type=COLLISION:NO INJURY Visibility='(6.75-inf)' Wind Speed='(7.333333-14.666667]' 799 ==> Precip='(-inf-0.005]' 792 conf:(0.99)
48. Visibility='(6.75-inf)' Wind Speed='(7.333333-14.666667]' 1797 ==> Precip='(-inf-0.005]' 1779 conf:(0.99)
49. Incident Type=COLLISION:NO INJURY Visibility='(6.75-inf)' Wind Speed='(-inf-7.333333]' 1097 ==> Precip='(-inf-0.005]' 1086 conf:(0.99)
50. Incident Type=COLLISION:NO INJURY Wind Speed='(7.333333-14.666667]' 817 ==> Precip='(-inf-0.005]' 808 conf:(0.99)

Apriori with collision incident attributes combined, using WEKA's mergeTwoValues filter:

1. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT & RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Sky Condition=CLR 2633 ==> Precip=(-inf-0.005] 2633 conf:(1)
2. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT & RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Sky Condition=CLR Visibility=(6.75-inf) 2628 ==> Precip=(-inf-0.005] 2628 conf:(1)
3. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT &

RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Status=IN PROGRESS Sky Condition=CLR 2367 ==> Precip=(-inf-0.005] 2367 conf:(1)

Apriori with collision incident attributes combined, using WEKA's mergeTwoValues filter (cont.):

4. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT & RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Status=IN PROGRESS Sky Condition=CLR Visibility=(6.75-inf) 2362 ==> Precip=(-inf-0.005] 2362 conf:(1)
5. Humidity=(42-71] 1908 ==> Precip=(-inf-0.005] 1908 conf:(1)
6. Humidity=(42-71] Visibility=(6.75-inf) 1905 ==> Precip=(-inf-0.005] 1905 conf:(1)
7. Humidity=(-inf-42] 1865 ==> Precip=(-inf-0.005] 1865 conf:(1)
8. Humidity=(-inf-42] 1865 ==> Visibility=(6.75-inf) 1865 conf:(1)
9. Humidity=(-inf-42] Visibility=(6.75-inf) 1865 ==> Precip=(-inf-0.005] 1865 conf:(1)
10. Precip=(-inf-0.005] Humidity=(-inf-42] 1865 ==> Visibility=(6.75-inf) 1865 conf:(1)
11. Humidity=(-inf-42] 1865 ==> Precip=(-inf-0.005] Visibility=(6.75-inf) 1865 conf:(1)
12. Sky Condition=CLR Humidity=(-inf-42] 1660 ==> Precip=(-inf-0.005] 1660 conf:(1)
13. Sky Condition=CLR Humidity=(-inf-42] 1660 ==> Visibility=(6.75-inf) 1660 conf:(1)
14. Sky Condition=CLR Humidity=(-inf-42] Visibility=(6.75-inf) 1660 ==> Precip=(-inf-0.005] 1660 conf:(1)
15. Sky Condition=CLR Precip=(-inf-0.005] Humidity=(-inf-42] 1660 ==> Visibility=(6.75-inf) 1660 conf:(1)
16. Sky Condition=CLR Humidity=(-inf-42] 1660 ==> Precip=(-inf-0.005] Visibility=(6.75-inf) 1660 conf:(1)
17. Status=IN PROGRESS Humidity=(42-71] 1584 ==> Precip=(-inf-0.005] 1584 conf:(1)
18. Status=IN PROGRESS Humidity=(42-71] Visibility=(6.75-inf) 1581 ==> Precip=(-inf-0.005] 1581 conf:(1)
19. Status=IN PROGRESS Humidity=(-inf-42] 1540 ==> Precip=(-inf-0.005] 1540 conf:(1)
20. Status=IN PROGRESS Humidity=(-inf-42] 1540 ==> Visibility=(6.75-inf) 1540 conf:(1)
21. Status=IN PROGRESS Humidity=(-inf-42] Visibility=(6.75-inf) 1540 ==> Precip=(-inf-0.005] 1540 conf:(1)
22. Status=IN PROGRESS Precip=(-inf-0.005] Humidity=(-inf-42] 1540 ==> Visibility=(6.75-inf) 1540 conf:(1)
23. Status=IN PROGRESS Humidity=(-inf-42] 1540 ==> Precip=(-inf-0.005] Visibility=(6.75-inf) 1540 conf:(1)
24. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT & RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Sky Condition=CLR Wind Speed=(-inf-7.333333] 1529 ==> Precip=(-inf-0.005] 1529 conf:(1)
25. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT & RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Sky Condition=CLR Visibility=(6.75-inf) Wind Speed=(-inf-7.333333] 1524 ==> Precip=(-inf-0.005] 1524 conf:(1)
26. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT & RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Humidity=(-inf-42] 1462 ==> Precip=(-inf-0.005] 1462 conf:(1)
27. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT & RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Humidity=(-inf-42] 1462 ==> Visibility=(6.75-inf) 1462 conf:(1)
28. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT & RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Humidity=(-inf-42] Visibility=(6.75-inf) 1462 ==> Precip=(-inf-0.005] 1462 conf:(1)
29. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT & RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Precip=(-inf-0.005] Humidity=(-inf-42] 1462 ==> Visibility=(6.75-inf) 1462 conf:(1)
30. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT &

RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Humidity=(-inf-42] 1462 ==> Precip=(-inf-0.005] Visibility=(6.75-inf) 1462 conf:(1)
31. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT & RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Humidity=(42-71] 1454 ==> Precip=(-inf-0.005] 1454 conf:(1)

Apriori with collision incident attributes combined, using WEKA's mergeTwoValues filter (cont.):

32. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT & RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Humidity=(42-71] Visibility=(6.75-inf) 1451 ==> Precip=(-inf-0.005] 1451 conf:(1)
33. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT & RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Status=IN PROGRESS Sky Condition=CLR Wind Speed=(-inf-7.333333] 1374 ==> Precip=(-inf-0.005] 1374 conf:(1)
34. Status=IN PROGRESS Sky Condition=CLR Humidity=(-inf-42] 1372 ==> Precip=(-inf-0.005] 1372 conf:(1)
35. Status=IN PROGRESS Sky Condition=CLR Humidity=(-inf-42] 1372 ==> Visibility=(6.75-inf) 1372 conf:(1)
36. Status=IN PROGRESS Sky Condition=CLR Humidity=(-inf-42] Visibility=(6.75-inf) 1372 ==> Precip=(-inf-0.005] 1372 conf:(1)
37. Status=IN PROGRESS Sky Condition=CLR Precip=(-inf-0.005] Humidity=(-inf-42] 1372 ==> Visibility=(6.75-inf) 1372 conf:(1)
38. Status=IN PROGRESS Sky Condition=CLR Humidity=(-inf-42] 1372 ==> Precip=(-inf-0.005] Visibility=(6.75-inf) 1372 conf:(1)
39. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT & RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Status=IN PROGRESS Sky Condition=CLR Visibility=(6.75-inf) Wind Speed=(-inf-7.333333] 1369 ==> Precip=(-inf-0.005] 1369 conf:(1)
40. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT & RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Status=IN PROGRESS Humidity=(-inf-42] 1313 ==> Precip=(-inf-0.005] 1313 conf:(1)
41. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT & RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Status=IN PROGRESS Humidity=(-inf-42] 1313 ==> Visibility=(6.75-inf) 1313 conf:(1)
42. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT & RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Status=IN PROGRESS Humidity=(-inf-42] Visibility=(6.75-inf) 1313 ==> Precip=(-inf-0.005] 1313 conf:(1)
43. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT & RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Status=IN PROGRESS Precip=(-inf-0.005] Humidity=(-inf-42] 1313 ==> Visibility=(6.75-inf) 1313 conf:(1)
44. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT & RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Status=IN PROGRESS Humidity=(-inf-42] 1313 ==> Precip=(-inf-0.005] Visibility=(6.75-inf) 1313 conf:(1)
45. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT & RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Sky Condition=CLR Humidity=(-inf-42] 1296 ==> Precip=(-inf-0.005] 1296 conf:(1)
46. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT &

RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Sky Condition=CLR Humidity=(-inf-42] 1296 ==> Visibility=(6.75-inf) 1296 conf:(1)

47. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT & RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Sky Condition=CLR Humidity=(-inf-42] Visibility=(6.75-inf) 1296 ==> Precip=(-inf-0.005] 1296 conf:(1)

Apriori with collision incident attributes combined, using WEKA's mergeTwoValues filter (cont.):

48. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT & RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Sky Condition=CLR Precip=(-inf-0.005] Humidity=(-inf-42] 1296 ==> Visibility=(6.75-inf) 1296 conf:(1)

49. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT & RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Sky Condition=CLR Humidity=(-inf-42] 1296 ==> Precip=(-inf-0.005] Visibility=(6.75-inf) 1296 conf:(1)

50. Incident Type=COLLISION:NO INJURY_HIT & RUN:PRIV PROP_HIT & RUN:NO INJURY_COLLISION:PRIV PROP_COLLISION:INJURIES_COLLISION:VS ANIMAL_COLLISION:VS DEER_HIT & RUN:INJURIES_COLLISION:NO DETAILS_HIT & RUN:UNKNOWN_COLLISION:FATALITY Status=IN PROGRESS Humidity=(42-71] 1295 ==> Precip=(-inf-0.005] 1295 conf:(1)