Resultant

Combining Geospatial Tools and Machine Learning to Map Predicted Crash Locations

After a crash, seconds can make the difference between life and death. A geospatial mapping solution not only enables citizens to avoid routes with high probability of a crash but helps police officers position themselves in those high-risk areas so that when a crash happens, they're close by and ready to assist. The result is fewer crashes and less chance that a crash becomes fatal.

CONTENTS

BACKGROUND

Imagine you're a state trooper and a vehicle collision has just occurred in your area. You need to get to the location as quickly as possible, attend to injured motorists, clear the road, and get traffic moving again to prevent additional crashes. Where would you position your patrol car to minimize response time?

You could use your gut and head to "that one spot" where experience tells you there might be a crash. But what if you could pair your intuition with a familiar, easy-to-use mapping tool that uses machine learning to predict areas where crashes are more likely to occur?

Resultant partnered with the Indiana State Police (ISP), the Indiana Department of Transportation (INDOT), and the Management Performance Hub (MPH) to create an interactive web-based map that provides law enforcement and the general public relevant information about where crashes have occurred in the past and where they are more likely to occur in the future.

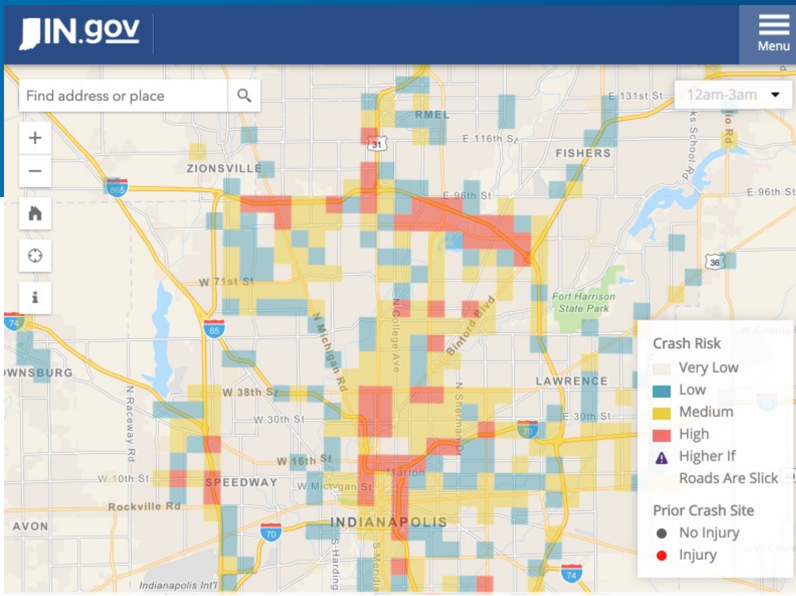


Figure 1: The Daily Crash Prediction Map shows the risk of a vehicle crash across the state throughout the day.

GATHERING DATA FROM MULTIPLE SOURCES

Data is the engine behind every data science project. Several data sources were brought together by Resultant and its partners to power the Daily Crash Prediction Map.

- The Automated Reporting Information Exchange System (ARIES) is a statewide crash repository with information about road classes, road surfaces, speed limits, and counts of previous crashes, all gathered from incident reports filled out by officers on scene.
- Average Annual Daily Traffic (AADT) volumes reveal traffic patterns on over 20,000 road segments through roadway sensors provided by INDOT. The patterns identify areas of high and low traffic flow on different days of the week during different times of day.
- Data was also collected from the U.S. Census and County Business Patterns (CBP). Information such as size of businesses, number of employers and employees, and population density within ZIP codes was used to determine locations with high commute-to-work rates and more overall traffic.
- Roads and crash locations were assigned to a grid of one-kilometer squares using the U.S. National Grid (USNG) reference system, which is used by first responders to more easily describe locations.

The data fed into the machine learning algorithm required significant processing. Road segments were clipped to the USNG polygons; latitudes and longitudes of every historical crash location were mapped to a grid using a point-in-polygon technique. Attributes of roads and previous crashes were then summarized for each grid.

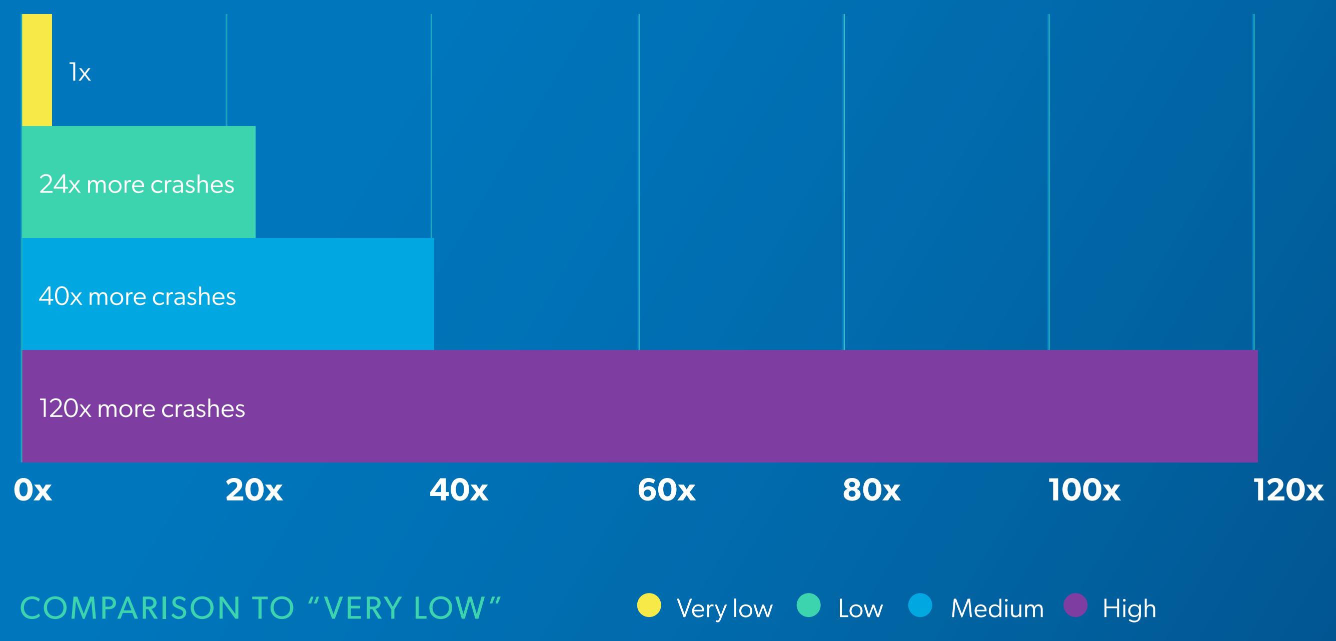
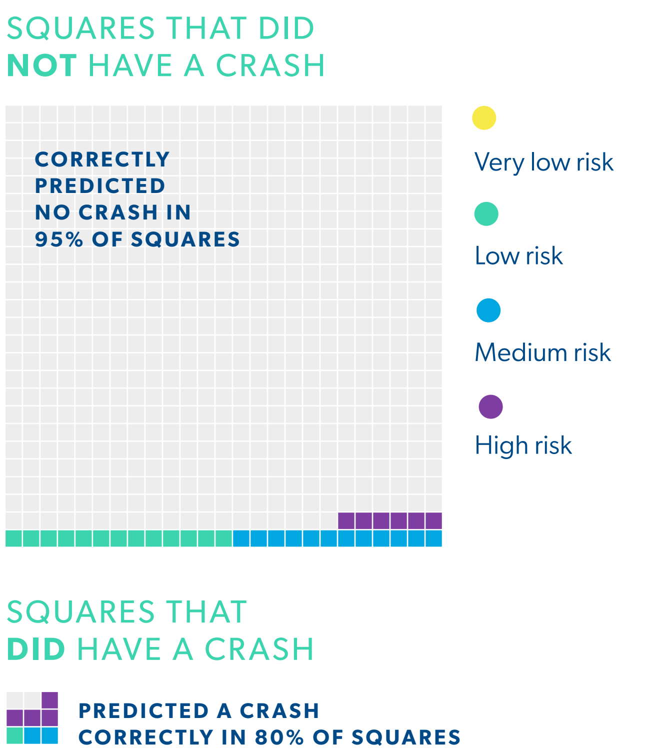


Figure 2: On average, areas predicted to be high risk will experience crashes >100 times more often than very low-risk areas. the day.

Geospatial Tools **in Action**

The Daily Crash Prediction Map uses data from the past to predict the future. The entire state is divided into one-kilometer squares. The map color codes each square based on the probability of a crash occurring during a three-hour period in the future, similar to a weather map. Users can search for locations, zoom in and out, pan, and change the time range.

Each day there are about 500,000 predictions across the state and around 460 crashes; relatively speaking, crashes are rare. Crashes occur in high-risk areas over 100 times more often than very low-risk areas. The tool's predictions were highly accurate: It correctly predicted crashes that did occur 80% of the time and correctly predicted where and when a crash would not occur 95% of the time.



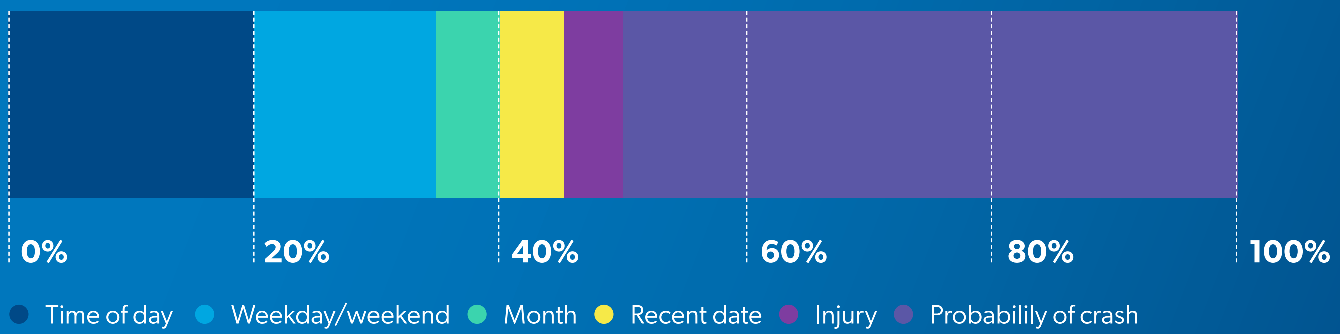


Figure 4: Resultant developed an algorithm to weight several attributes that determine which prior crashes to display.

Solving Unique **Challenges**

Millions of vehicles have crashed in Indiana since 2004, and police collect a wealth of data for each collision. This presents challenges in executing data analysis, ensuring the resulting analysis is useful, and protecting privacy.

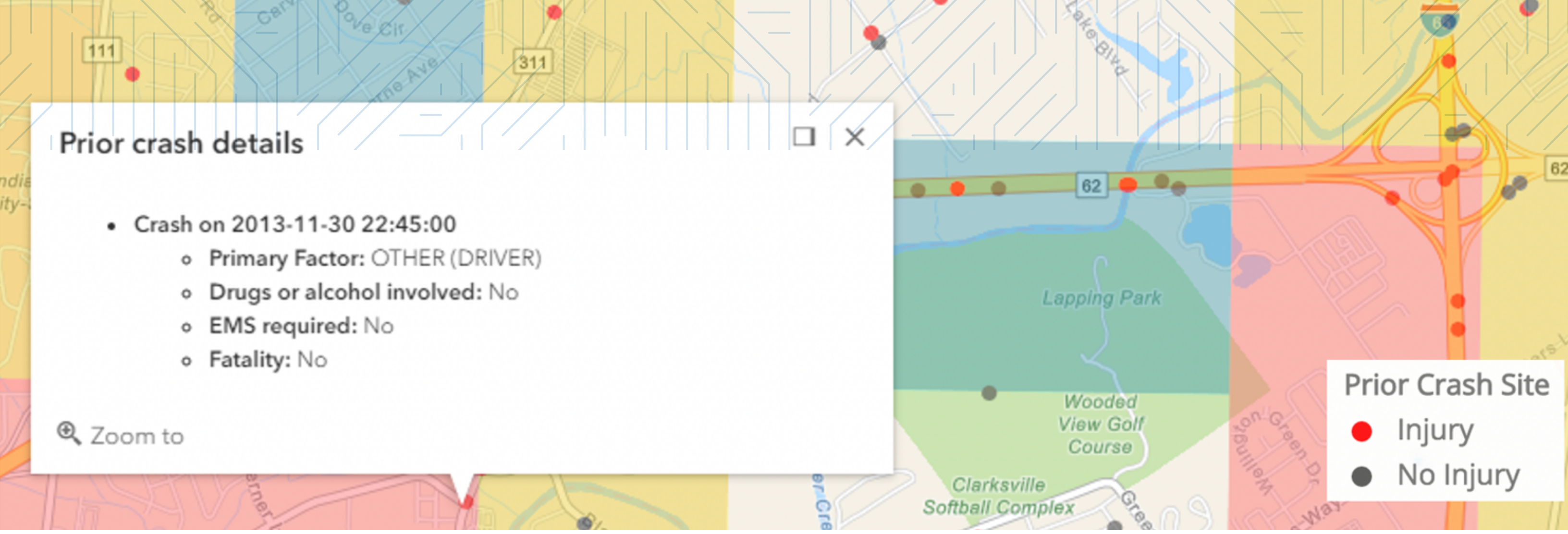
PROVIDING INSIGHT RATHER THAN INFORMATION OVERLOAD

Map users said that seeing markers showing locations where previous crashes took place would be helpful. Yet displaying millions of points at once on a map is a recipe for user frustration. In such a case, the map would be slow to load, pan, and zoom, and the user wouldn't be able to comprehend the immense amount of information.

We listened deeply to users to discover which characteristics of the previous crashes were most important to them. Then Resultant data scientists developed an algorithm that ranked each crash, identifying the most relevant and representative ones. For instance, if the current time is 3pm on a Tuesday, the algorithm is more likely to show crashes around that time of day from previous Tuesdays rather than crashes on Sunday at 2am. Similarly, crash patterns are different in January than July, recent crashes are more relevant than crashes from five years ago, and it's important to see crashes that resulted in injury or death. The algorithm highlights crashes based on each of these attributes; it also displays more historical crashes in areas where the probability of a future crash is higher.

In addition to limiting the number of incidents displayed, the map shows prior crashes only after the user zooms in to a particular area. Dots representing crashes are color coded to indicate whether they resulted in injury. The user can click on each crash and see several attributes: date and time, primary factor behind the crash, whether drugs or alcohol were involved, whether an ambulance was called, and whether a fatality occurred. The resulting map is easy to use and provides the right level of information at the right time.

Figure 5: Crashes have happened pretty much everywhere along IN roads—good luck finding dangerous locations from a map like this!



Dealing with **Lopsided Data**

A machine learning algorithm that predicts whether something will happen or not is called a classification model. The model learns by analyzing the same type of data for each prediction class. In this example, the classes are whether a crash will occur ("crash") or not ("no crash") in a specific area, during a specific time window.

This crash data is lopsided in two respects. The first is that crashes are thankfully rare—there are orders of magnitude more instances of the no-crash class than the crash class. This is known as a class imbalance, and Resultant data scientists used techniques like undersampling to alleviate the issue.

The more challenging issue is a result of how crash data is collected. When a crash occurs, law enforcement records a wealth of useful data. When a crash does not occur, though, we have no specific information about the drivers, vehicles, or exact road conditions. We know how many crashes involved a drowsy truck driver hitting a patch of black ice while distracted by a cell phone, but we have no similar information about times when a crash did not happen; the data is one-sided.

Resultant data scientists and subject matter experts got creative to overcome this challenge. They developed several features of data that could be known ahead of time and would be available for both crash and no crash instances. Some of the features used to train the machine learning model included:

- Road composition: asphalt, cement, or other
- Number of lanes on the road
- Presence of a center divider
- Speed limits
- Complexity of the road network—from a single two-lane road to several high-traffic roads and major limited access highways
- Active construction or repair work
- General weather conditions
- Angle of the sun
- Holidays
- Number and size of nearby businesses
- Frequency of previous crashes at similar times

Preserving **Geospatial Privacy**

Data privacy was not a primary concern for this crash prediction project, as the information provided by the tool does not allow reidentification of individual-level data, and collisions are a matter of public record. For other geospatial applications, though, such as the Naloxone Administration Heatmap, privacy must be preserved while still revealing relevant location information. For that project, Resultant helped MPH create a map for the Indiana Department of Homeland Security (IDHS) that reveals the approximate areas where naloxone was administered to reverse an opioid overdose without revealing the address. First responders can get a better idea of where naloxone has been administered and plan accordingly; public health officials can target programs where they're needed most.

Resultant data scientists developed a geospatial masking algorithm to preserve the privacy of individuals receiving naloxone. The latitude and longitude of each point is displaced a random distance in a random direction.

The distance is shorter in ZIP codes with high population and housing density, and longer in sparser areas. Where multiple incidents occur at the same location, all points are initially displaced the same distance in the same direction, and then each point is displaced an additional random (but shorter) distance in a random direction. Without this, the average of a cluster of points would be very close to the unmasked original location.

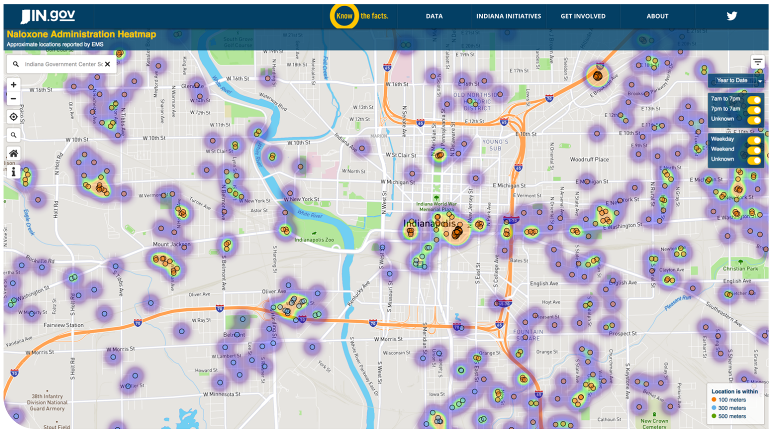


Figure 7: The Naloxone Administration Heatmap preserves location privacy while revealing neighborhood trends.

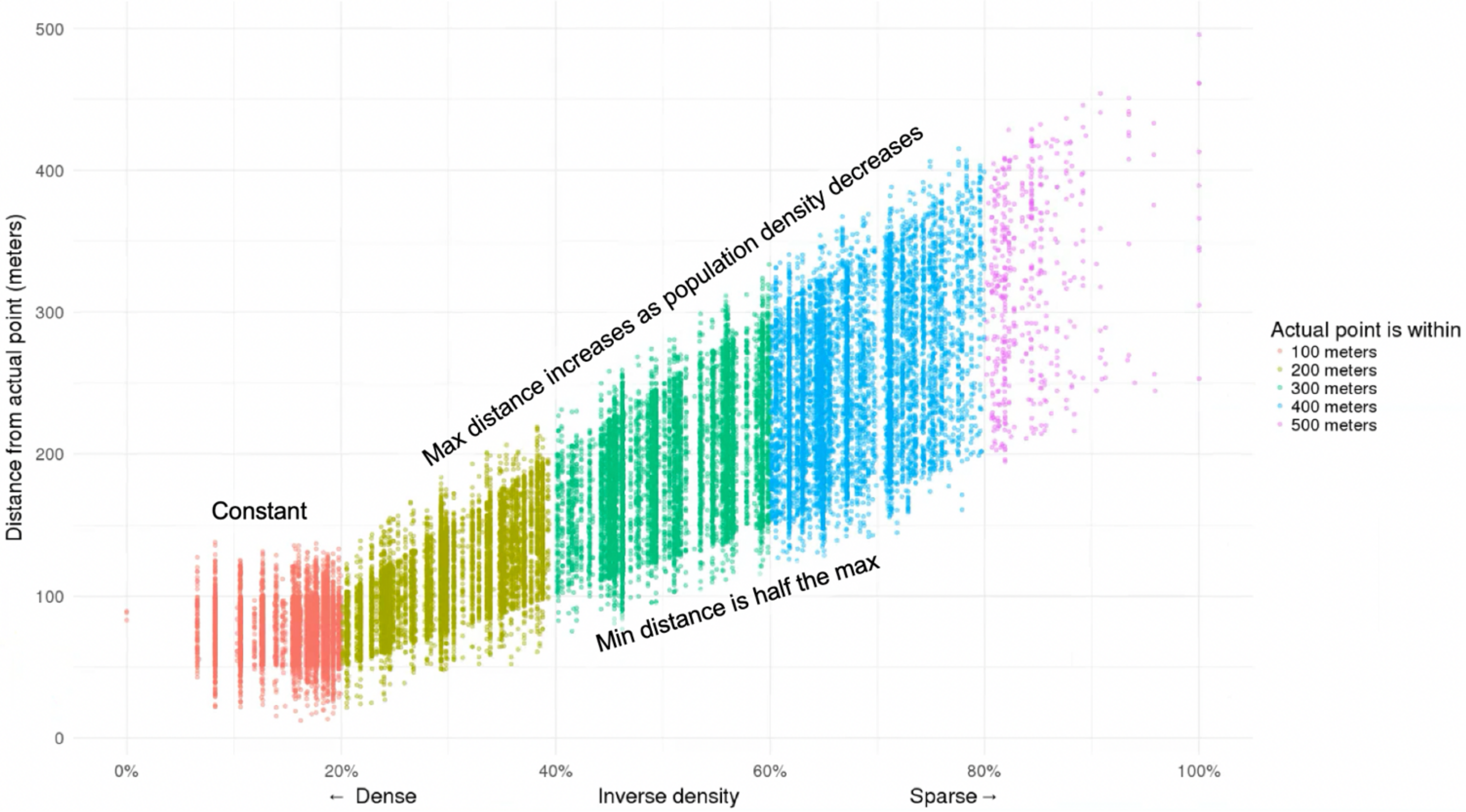


Figure 8: The Resultant masking algorithm displaces points a random distance that increases with decreasing population density.

Putting the Solution **into Production**

Some data science projects provide an initial insight, and then the static results are set aside. Since the Daily Crash Prediction Map would be used daily by state police and the general public, it needed to be available around the clock with refreshed historical crash information and up-to-date crash predictions.

Resultant worked alongside MPH to build a fully automated solution for ingesting new data: preparing it, creating predictions, and uploading the results to the mapping tool. Resultant also built the web-based interactive mapping application. The tool was in production for several years, and there were almost no cases where the pipeline or tool failed—it basically ran itself.

IMPACT

Using our solution, officers can better position themselves in high-risk areas, cutting down response times to crash sites. In a separate analysis, Resultant found that quick responses to a crash involving an injury are associated with a reduced likelihood of the injury being fatal. Seconds can make the difference between death and life; being close to the crash site when it happens matters.

The faster an officer can get there, the faster they can get traffic flowing again and prevent another collision. When traffic backs up because of the primary crash, there's an increased risk for a secondary crash. Full-speed vehicles suddenly colliding with stopped cars can create even more severe injuries than the original accident.

Finally, the tool enables citizens to take a proactive role in traffic safety. They can adjust their routes or be more conscious of risk in areas with high probability of a crash, leading to fewer crashes.

OTHER USES OF GEOSPATIAL MAPPING

Resultant's geospatial mapping techniques could have many applications. In addition to turning millions of points into actionable insight and displaying spatiotemporal results of machine learning models in a user-friendly map, real-time traffic, weather forecasts, and image recognition can be incorporated to accurately predict various outcomes. Resultant has a strong track record of developing custom, novel solutions to clients' most challenging and complex problems.

ABOUT RESULTANT

Our team believes solutions are more valuable, transformative, and meaningful when reached together. Through outcomes built on solutions rooted in data analytics, technology, and digital transformation, Resultant serves as a true partner by solving problems with our clients, rather than for them.

DATA ANALYTICS

We help organizations understand their data landscape and solve problems by turning data into insight. While data can be dense, our team's empathetic approach to problem solving creates meaningful solutions with deep technical foundations.

© 2021 Resultant