

# Big Data Applications for Vehicle Crash Prediction

Kevin Duffy  
Indiana University  
4014 E. Stop 10 Rd.  
Indianapolis, Indiana 46237  
kevduffy@iu.edu

## ABSTRACT

The idea of predicting car crashes used to be a fruitless endeavor - relegated to mere guesswork. However, advances in big data applications have allowed law enforcement to more accurately predict when and where car crashes are likely to occur, thus lowering first response times and taking proactive actions to prevent accidents in high-risk corridors. This paper shows several different approaches different agencies have taken using the power of data to solve this critical problem.

## KEYWORDS

Big Data, Vehicles, Crashes, HID310, I523

## 1 INTRODUCTION

Car crashes are the top cause of death for Americans between ages 5 and 34.[8] Although the rate of car crashes has been dropping steadily over the last few decades [2], they remain a constant hazard for any who travel on our roads.

While most of the talk regarding car crashes and big data usually involves the development of self-driving cars (and understandably so), more immediate measures are being developed to meet this problem. A common usage of data in this domain is the prediction of high-risk crash areas.

The idea of predicting when and where crashes would occur used to be pure guesswork on the part of experienced police troopers[5], but advances in big data analysis has allowed law enforcement agencies to begin predicting and mapping exactly when and where high-risk areas will occur. Advocates contend that these tools allow first responders to act more efficiently both reactively and proactively, making our roads safer on which to travel.

We will outline the initial problem being addressed by these tools, the ultimate goal of such an exercise, and briefly explore two approaches to this solution. We will determine what outcomes these applications had, as well as whether they can be fully evaluated at this time.

## 2 TRENDS AND GOALS

In 1975, the United States had a rate of 3.35 deaths per 100 million miles traveled.[2] In 2015, that rate stood at 1.13 deaths. In fact, during that time frame, the amount of miles traveled on our roads has more than doubled, while the average annual number of fatalities on our roads has dropped by more than 10,000. The National Highway Traffic Safety Administration (NHTSA) credits this to the successful implementation of the Four Es of traffic safety:

- Enforcement
- Engineering
- Education

- Emergency Medical Services [8]

However, the NHTSA and other safety groups are not satisfied with merely lower numbers. They are driven by an initiative called Toward Zero Deaths which, as the name implies, strives toward the elimination of traffic related deaths.[8]

But how can highway safety groups reach this lofty goal? And what are they already doing about it? It should come as little surprise that the powerful utilization of big data has been rising in this field.

## 3 BIG DATA SOLUTIONS

A lot of data points can be collected from current activities undertaken through the Four Es. These include things such as traffic volume at the time of the crash, vehicle speed, road and construction conditions, and emergency response times. [8]

In addition, state agencies have found that other extraneous factors contributed to the likelihood of a crash, such as events like football games, major holidays that correspond with an increase in drunk driving (such as New Year's Eve and Super Bowl Sunday), and concentrations of establishments that serve alcohol.[5]

Unfortunately, these data points were previously siloed off in different agencies and formats, making dynamic usage of this information difficult if not impossible. However, some states are beginning to utilize this data in an effective way. They are beginning to collect data in a deliberate way so as to coordinate information between agencies in order to create applications and tools for use by troopers, other emergency response, road planners, and the public at large. [5]

States are going about this problem in different ways. This will allow us to approach the issue from different angles and see which approaches yield superior results. However, since these applications are still so new, it is not yet possible to fully evaluate their effectiveness on outcomes. This paper is examining two state approaches: Tennessee, which was one of the first initiatives in the nation, and Indiana, which was modeled after Tennessee's program in a more comprehensive way.

### 3.1 Tennessee

The origin of Tennessee Highway Patrol's Predictive Analytics program began in 2008, when state troopers switched from paper-and-pen crash reports to an electronic interface. This allowed Tennessee's agencies to use real-time data to create prediction software for car crashes. According to the State of Tennessee, the program uses SPSS software to apply three different statistical models with machine-learning algorithms in order to provide troopers with risk areas for certain types of crashes. [4] The models used were:

- (1) Crash Reduction Analyzing Statistical History (CRASH). Using risk factors and historical crash data, this model gives probabilities on a fatal or injury-causing accident over four-hour blocks over a one week period, which is then illustrated on interactive maps.
- (2) Driving Under the Influence (DUI). Calculates the probability of a crash related to a DUI from 4 pm to 4 am.
- (3) Commercial Motor Vehicle (CMV). Calculates the probability of a crash related to a commercial motor vehicle (such as a semi truck).[4]

The program was implemented in 2013. Using this technology, troopers stationed themselves in more statistically advantageous positions in order to decrease response times to crashes, and take preliminary actions to prevent crashes such as setting up traffic direction or more aggressive enforcement on speeding and reckless driving.

From 2013 to 2015, Tennessee traffic deaths fell 3 percent, compared to a 7 percent increase across the nation. This cannot yet be directly tied to the performance of the program, but according to officials such as the THP Statistics official Patrick Dolan, the effect is clear.

Information coming out of the Predictive Analytics program driving targeted enforcement "has allowed us to deploy our resources more effectively to execute our mission successfully," Dolan reported to the Tennessee government's Traffic Safety Innovations 2016 newsletter. [4]

However, the trend became murkier in 2016. Tennessee traffic deaths spiked 8 percent, matching the national trend. Officials are still unclear whether the model is to blame. According to Dolan, average police response time dropped nearly 33 percent between 2012 and 2016. [5]

### 3.2 Indiana

In March 2014, then-Indiana Governor Mike Pence create by executive order the Management Performance Hub (MPH), a subagency tasked with driving "a coordinated effort by state agencies to share data and improve and strengthen services, maximize the utilization of available resources, and ensure that state services are available to all Hoosiers." [3]

One of the first projects undertaken by MPH was the Crash Prediction Website. Inspired by Tennessee's program [7], the MPH worked in tandem with the Indiana State Police to create an interactive map (Figure 1) showing the probability of both fatal and nonfatal traffic accidents.

[Figure 1 about here.]

The model grew out of several factors:

- (1) The concept was borrowed from Tennessee's aforementioned Predictive Analytics program, though Indiana had a greater range and depth of data available.
- (2) The core of the model is built upon data from 2 million crashes going back to 2004. The data was cleaned so only relevant crash data was included in the forecast.
- (3) The probability of a car crash is ranked from "very low risk" to "high risk", which is then indicated on the map through color-coded 1 square kilometer blocks over three hour time blocks.[6] This is more granular than Tennessee's

maps, in both time and space scale. These probabilities are ranked based off of weather forecasts, traffic congestion, road conditions, time of day, historical information, and census data.

- (4) The map is populated with different colored dots to represent past car crashes - red for injury-causing, grey for non-injuries. Users can then click these to identify unique details of actual car crashes.[7]
- (5) The model is updated automatically with fresh crash reports, providing the software with dynamic information.

Unlike Tennessee's program, Indiana's crash map is completely available to the public for their own personal use, though its usage will primarily be used by police and first responders.

Although it is still too soon after launch to evaluate the effectiveness of this model, Indiana officials are optimistic. Indiana's Office of Management and Budget, the parent agency to the MPH, estimates that even a one percent reduction in Indiana crashes could net up to \$30 million in annual savings, besides the obvious benefits of fewer road casualties.

With the groundwork laid in the crash map, officials are hoping to utilize the technology in other ways. Major Michael White of the Indiana State Police, in an interview with Statescoop.com, said with the technology developed through this initiative, they hope to move on to using it to map crime data as well.[6]

## 4 CONCLUSION

While these applications and initiatives are still too new to fully evaluate, there does appear to be preliminary results that show promise.

Tennessee showed quick decline in car crashes and police response times, while Indiana built upon Tennessee's example to provide a more comprehensive look at all risk factors involved in a car crash, while opening this tool up to the public for personal use.

These applications also show promise in application to other domains, such as the Indiana State Police's interest in creating a crime risk map using the same principles.

It is also encouraging to view these as an example in various state agencies coordinating in order to share data and collaborate on applications. If Indiana's Management Performance Hub is any indication, these collaborations will continue, at least in Indiana. Hopefully more states follow suit in creating data sharing hubs and protocols in order to streamline government data utilization.

Ultimately, what is primarily needed is more time to evaluate the effectiveness of these initiatives, as well as metadata related to utilization of the map (for example, what is the average response times of a trooper using the map versus one who is not using the map). As more states grow their own initiatives, we can also evaluate whether certain approaches are more effective than others.

## ACKNOWLEDGMENTS

The author would like to thank Dr. Gregor von Laszewski and all TA's for their tireless work in ensuring this class goes smoothly.

## REFERENCES

- [1] [n. d.]. ([n. d.]). <https://www.in.gov/isp/ispCrashApp/main.html>
- [2] [n. d.]. General statistics. ([n. d.]). <http://www.ihs.org/ihs/topics/t/general-statistics/fatalityfacts/overview-of-fatality-facts>

- [3] 2014. . Number 14-06. 2 pages.
- [4] 2016. *State of Tennessee* (2016). <https://www.tn.gov/assets/entities/safety/attachments/Technology.v1.pdf>
- [5] Jenni Bergal. 2017. Troopers Use 'Big Data' to Predict Crash Sites. (Feb 2017). <http://www.pewtrusts.org/en/research-and-analysis/blogs/stateline/2017/02/09/troopers-use-big-data-to-predict-crash-sites>
- [6] Mackenzie Carmen. 2016. Indiana State Police unveil Daily Crash Prediction Map. (Nov 2016). <http://statescoop.com/indiana-state-police-unveil-car-crash-forecast-map-to-help-reduce-traffic-accidents-in->
- [7] Eyrageon Eidam. 2016. Indiana Launches Predictive Crash Tool for Citizens, First Responders. (Nov 2016). <http://www.govtech.com/data/Indiana-Launches-Predictive-Crash-Tool-for-Citizens-First-Responders.html>
- [8] Melissa Savage. 2012. Crash analytics: How data can help eliminate highway deaths. (Apr 2012). <https://gcn.com/articles/2012/04/20/data-analytics-traffic-safety-toward-zero-deaths.aspx>

## A ISSUES

DONE:

Example of done item: Once you fix an item, change TODO to DONE

### A.1 Assignment Submission Issues

DONE:

Do not make changes to your paper during grading, when your repository should be frozen.

### A.2 Uncaught Bibliography Errors

DONE:

Missing bibliography file generated by JabRef

DONE:

Bibtex labels cannot have any spaces, \_ or & in it

DONE:

Citations in text showing as [?]: this means either your report.bib is not up-to-date or there is a spelling error in the label of the item you want to cite, either in report.bib or in report.tex

### A.3 Formatting

DONE:

Incorrect number of keywords or HID and i523 not included in the keywords

DONE:

Other formatting issues

### A.4 Writing Errors

DONE:

Errors in title, e.g. capitalization

DONE:

Spelling errors

DONE:

Are you using *a* and *the* properly?

DONE:

Do not use phrases such as *shown in the Figure below*. Instead, use *as shown in Figure 3*, when referring to the 3rd figure

DONE:

Do not use the word *I* instead use *we* even if you are the sole author

DONE:

Do not use the phrase *In this paper/report we show* instead use *We show*. It is not important if this is a paper or a report and does not need to be mentioned

DONE:

If you want to say *and* do not use & but use the word *and*

DONE:

Use a space after . , :

DONE:

When using a section command, the section title is not written in all-caps as format does this for you

\section{Introduction} and NOT \section{INTRODUCTION}

### A.5 Citation Issues and Plagiarism

DONE:

It is your responsibility to make sure no plagiarism occurs. The instructions and resources were given in the class

DONE:

Claims made without citations provided

DONE:

Need to paraphrase long quotations (whole sentences or longer)

DONE:

Need to quote directly cited material

### A.6 Character Errors

DONE:

Erroneous use of quotation marks, i.e. use "quotes", instead of " " "

DONE:

To emphasize a word, use *emphasize* and not "quote"

DONE:

When using the characters & # % \_ put a backslash before them so that they show up correctly

DONE:

Pasting and copying from the Web often results in non-ASCII characters to be used in your text, please remove them and replace accordingly. This is the case for quotes, dashes and all the other special characters.

DONE:

If you see a ffigure and not a figure in text you copied from a text that has the fi combined as a single character

## A.7 Structural Issues

DONE:

Acknowledgement section missing

DONE:

Incorrect README file

DONE:

In case of a class and if you do a multi-author paper, you need to add an appendix describing who did what in the paper

DONE:

The paper has less than 2 pages of text, i.e. excluding images, tables and figures

DONE:

The paper has more than 6 pages of text, i.e. excluding images, tables and figures

DONE:

Do not artificially inflate your paper if you are below the page limit

## A.8 Details about the Figures and Tables

DONE:

Capitalization errors in referring to captions, e.g. Figure 1, Table 2

DONE:

Do use *label* and *ref* to automatically create figure numbers

DONE:

Wrong placement of figure caption. They should be on the bottom of the figure

DONE:

Wrong placement of table caption. They should be on the top of the table

DONE:

Images submitted incorrectly. They should be in native format, e.g. .graffle, .pptx, .png, .jpg

DONE:

Do not submit eps images. Instead, convert them to PDF

DONE:

The image files must be in a single directory named "images"

DONE:

In case there is a powerpoint in the submission, the image must be exported as PDF

DONE:

Make the figures large enough so we can read the details. If needed make the figure over two columns

DONE:

Do not worry about the figure placement if they are at a different location than you think. Figures are allowed to float. For this class, you should place all figures at the end of the report.

DONE:

In case you copied a figure from another paper you need to ask for copyright permission. In case of a class paper, you must include a reference to the original in the caption

DONE:

Remove any figure that is not referred to explicitly in the text (As shown in Figure ..)

DONE:

Do not use `textwidth` as a parameter for `includegraphics`

DONE:

Figures should be reasonably sized and often you just need to add `columnwidth`

e.g.

```
/includegraphics[width=\columnwidth]{images/myimage.pdf}  
re
```

## LIST OF FIGURES

- 1 The map shows the probability of a crash through color-coded blocks.[1]

6

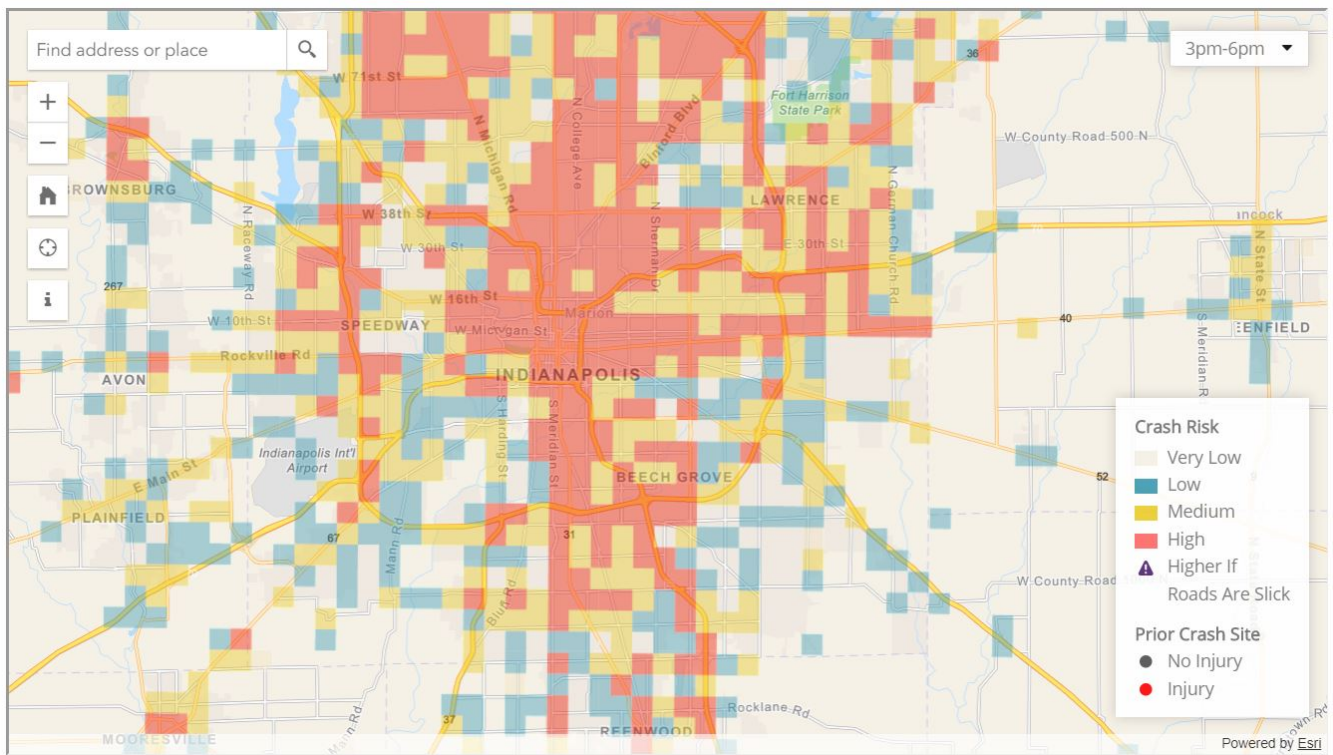


Figure 1: The map shows the probability of a crash through color-coded blocks.[1]