



Crash Predictive Analytics A Different Approach

CAPS

CENTER *for* ADVANCED
PUBLIC SAFETY

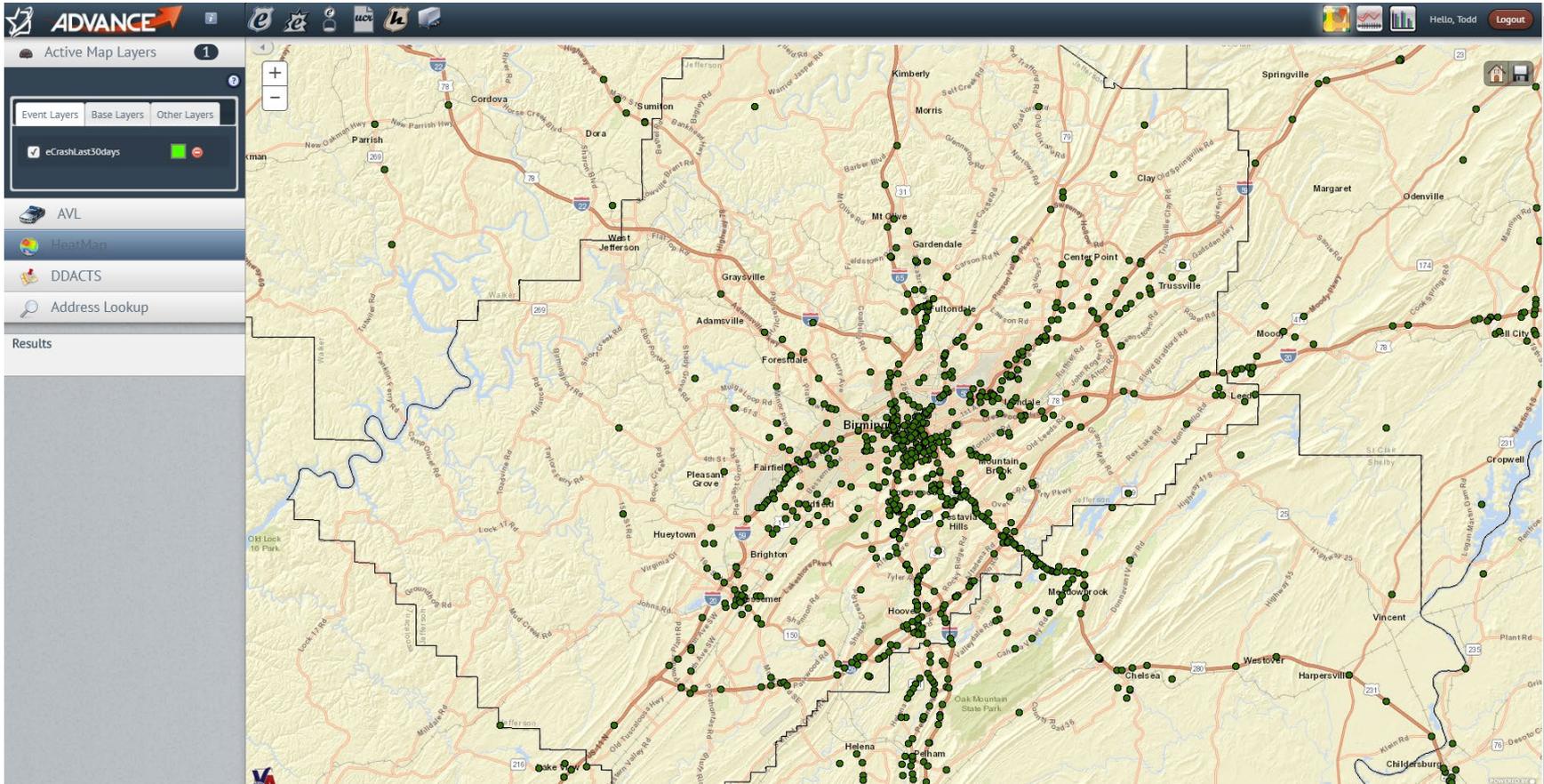
Traditional “Predictive Analytics”

- Assumes consistent system behavior
- If it happened here before at this time, it’ll happen again
- Leans heavily on accurate and comprehensive crash report data
- Does promote the requirement for good data collection, which is important no matter what you’re doing

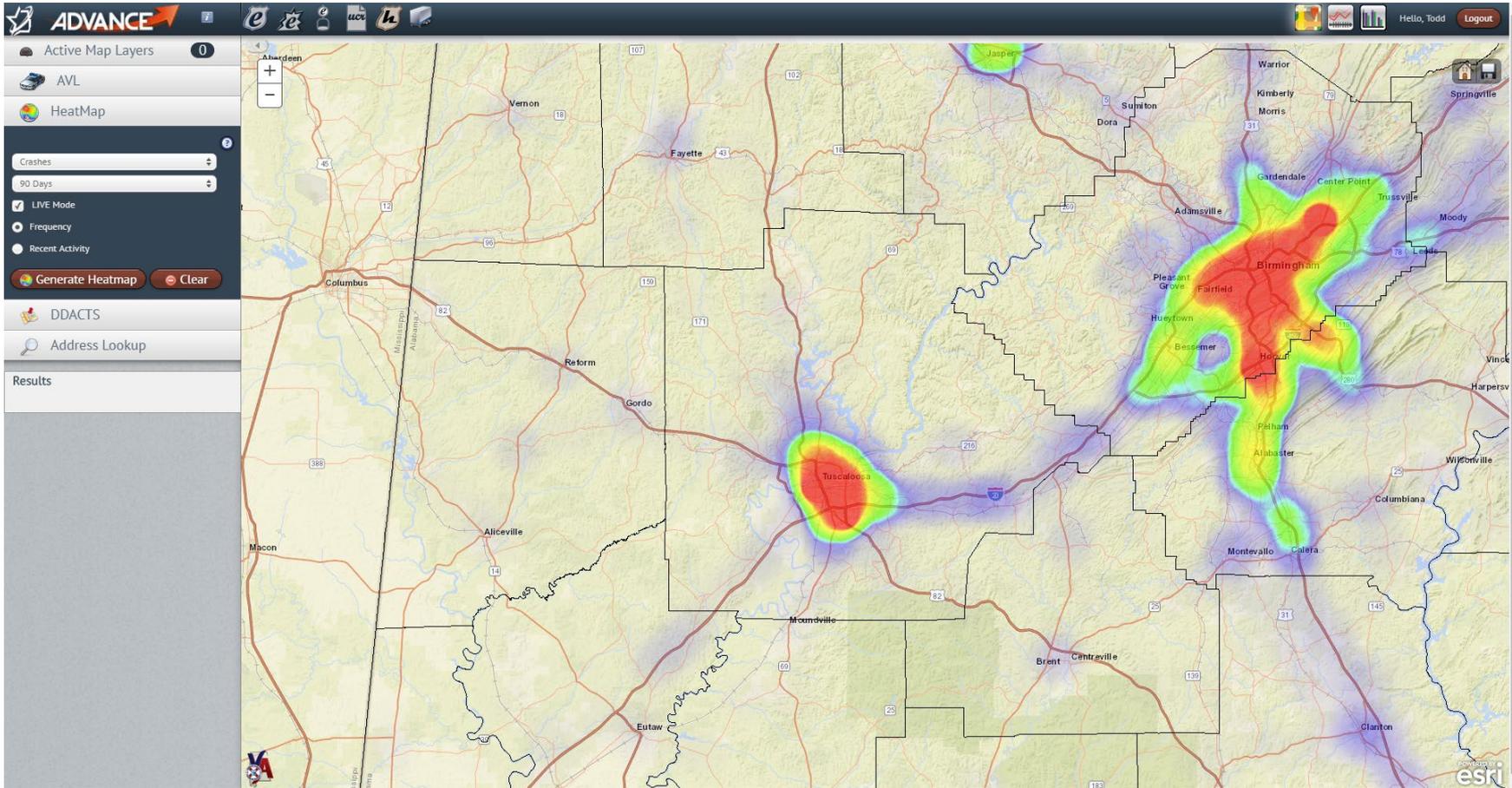
ADVANCE

- Multiple datasets (crash, citation, etc.)
- Custom map layers based on filters (e.g. CMV crashes in a specific county)
- Aggregate mapped hotspots based on historical data.
- Good situational and historical awareness
- Less desirable for predictions

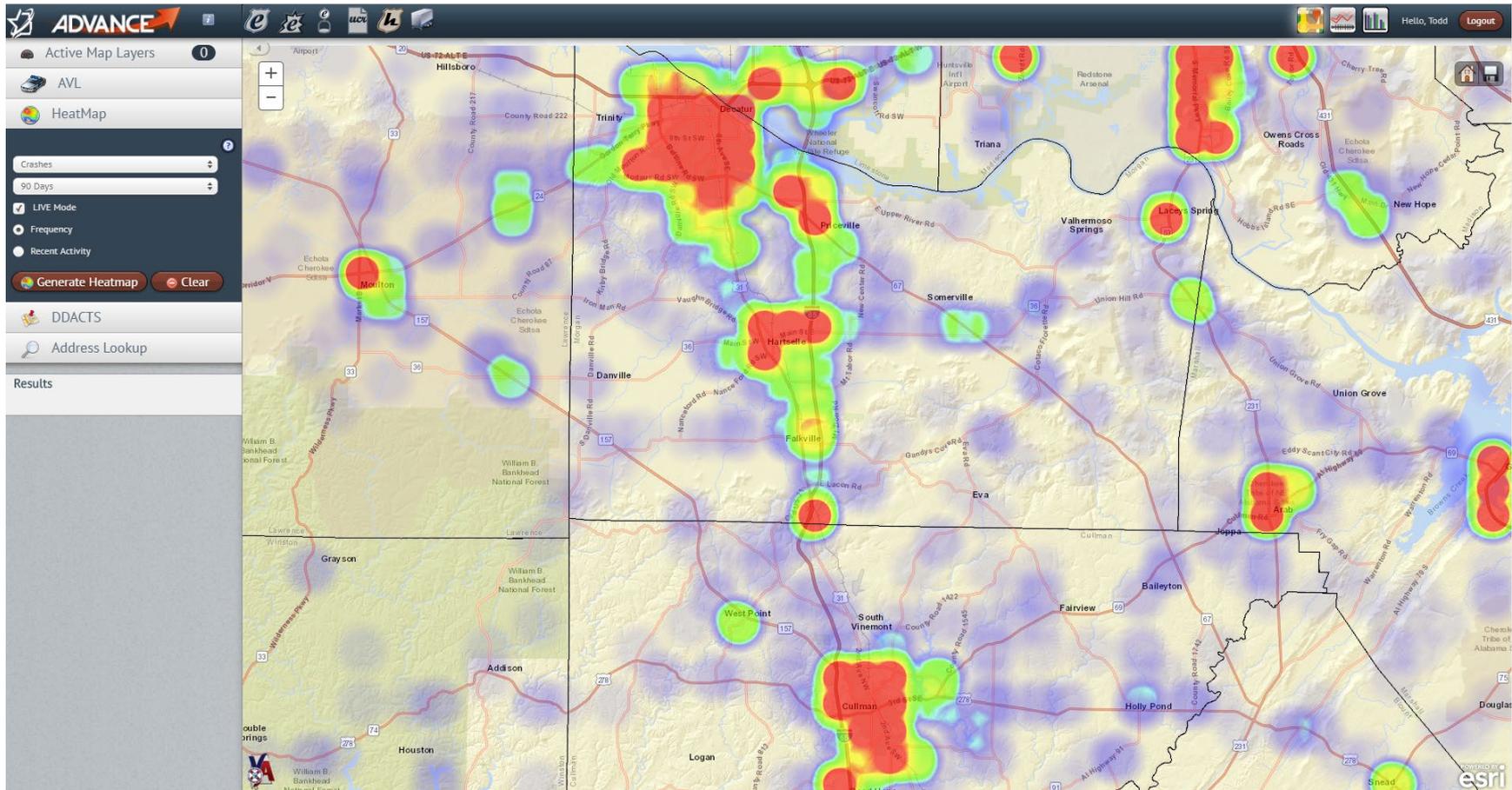
Custom Map Layer w/ Points



Focused Heat Mapping - Wide



Focused Heat Mapping - Zoomed



Chasing the Dream of Perfect Predictions

“People's environments change even more quickly than they themselves do. Everything from the weather to their relationship with their mother can change the way people think and act. All of those variables are unpredictable. How they will impact a person is even less predictable. If put in the exact same situation tomorrow, they may make a completely different decision. This means that a statistical prediction is only valid in sterile laboratory conditions, which suddenly isn't as useful as it seemed before.”

- Dr. Gary King

Current Methods

- Pretty good overall.
- Lots of sub-approaches, with pros/cons of each.
- Mostly regression-model based (some machine learning thrown in, also)
- No “silver bullet” method.

Things to Improve

1. Add more peripheral data...
2. Build user's trust (at all levels)...

Peripheral Data

- Does everything that contributed to the crash get documented on the report?
- How far does relevant data extend?
- More data, more accurate model (*mostly...*)
- Peripheral data may be useless (or at least doesn't contribute to crashes...)
- **Let the machine filter it for you.**

Peripheral Data

- Geographic (e.g. retail/commercial/industrial locations, roadway characteristics, etc.)
- Temporal (e.g. light level to time relationship)
- Behavioral (e.g. social media “sentiment”, top story news, etc.)
- Economic (e.g. GDP, unemployment, etc.)
- *Anything, really...*

Trusting the Predictions

The predictions don't always have to make sense to you, but it helps...

Trusting the Prediction



Trusting the Prediction



Trusting the Predictions

Promote

Proving that the model beats other methods. At least, it shouldn't make things worse.

Meshing the human with the prediction system in a balanced way.

Novel recommendations that make sense.

Avoid

Blind trust in the model.

Ignoring completely or trusting totally user contributed "danger zones".

"Novel" recommendations based on randomness or "dartboard predictions"

Our Approach: *CrashPredict*

- Automated inclusion of peripheral data.
- Parallel model analysis.
 - Did we do better than a standard deployment model?
- Regression-based for core analysis.
- Open-ended machine learning model for peripheral data and as a sandbox.

CrashPredict Goals

Goal 1: *Universal data import.*

Goal 2: *Tunable model, based on target.*

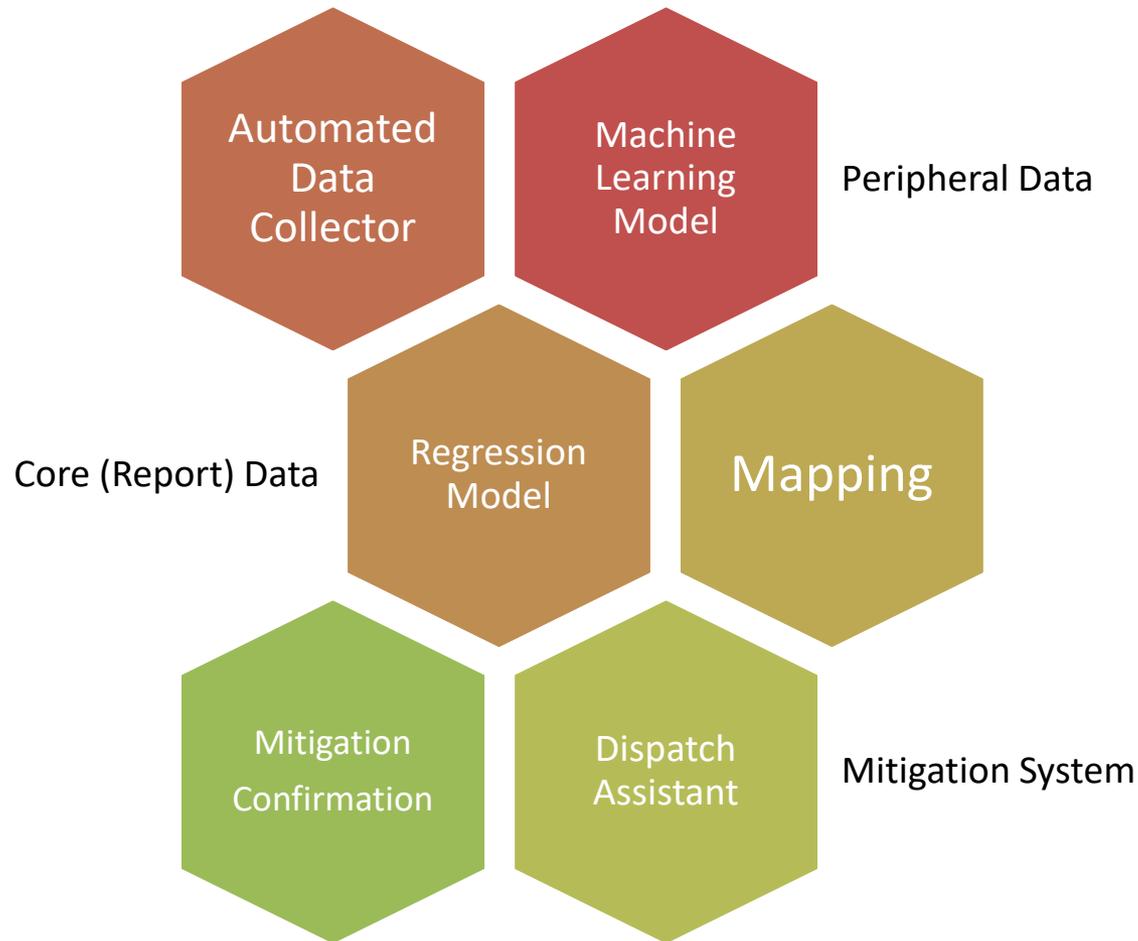
Goal 3: *Results focused system. Positive impact is everything.*

Goal 4: *Trust in the model. Model should make sense, at least to someone.*

CrashPredict Components

- Automated inclusion of peripheral data.
- Parallel model analysis.
 - Did we do better than a standard deployment model?
- Regression-based for core analysis.
- Open-ended machine learning model for peripheral data and as a sandbox.

CrashPredict Components





Thank You!

Jeremy Pate
Associate Director of Technology
jrpate@ua.edu
205-210-8043