THE UNIVERSITY OF WESTERN AUSTRALIA

School of Computer Science and Software Engineering

CITS4009 Introduction to Data Science

SEMESTER 2, 2017: PART 3 DELIVERING RESULTS

CHAPTER 11 PRODUCING EFFECTIVE PRESENTATIONS

Chapter Objectives

- Presenting your results to project sponsors
- Communicating with your model's end users
- Presenting your results to fellow data scientists

Example

Our company (let's call it WVCorp) makes and sells home electronic devices and associated software and apps. WVCorp wants to monitor topics on the company's product forums and discussion board to identify "about-to-buzz" issues: topics that are posed to generate a lot of interest and active discussion. This information can be used by product and marketing teams to proactively identify desired product features for future releases, and to quickly discover issues with existing product features. Once we've successfully built a model for identifying about-to-buzz topics on the forum, we'll want to explain the work to the project sponsor, and also to the product managers, marketing managers, and support engineering managers who will be using the results of our model.

Entities in the buzz model scenario

Entity	Description
WVCorp	The company you work for
eRead	WVCorp's e-book reader
TimeWrangler	WVCorp's time-management app
BookBits	A competitor's e-book reader
GCal	A third-party cloud-based calendar service that TimeWrangler can integrate with

Presenting your results to the project sponsor

1. Summarize the motivation behind the project, and its goals.

2. State the project's results.

3. Back up the results with details, as needed.

4. Discuss recommendations, outstanding issues, and possible future work.

- Some people also recommend an "Executive Summary" slide: a one-slide synopsis of steps 1 and 2.
- How you treat each step—how long, how much detail—depends on your audience and your situation.

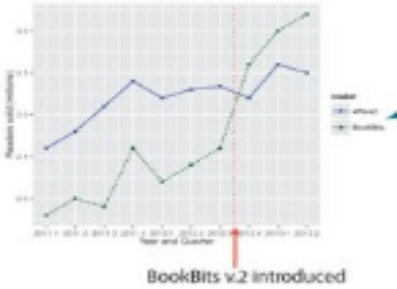
Summarizing the project's goals

1- Motivation for project

If Only We'd Known...

eRead vs. BookBits

- eRead: Best selling indie ebook reader - until BookBits v.2
- Hot new BookBits feature: shared bookshelves
  - Our one-at-a-time book lending didn't compete
- Estimated $25M lost revenue on product sales



Show the business need that motivated this project: WVCorp lost revenue and market share to a competitor.

Solid curve: WVCorp's product (eRead)
Dashed curve: Competitor's product (BookBits)

Could We Have Caught This?

- eRead Forum discussions:
  - Sharing a booklist with a friend, to grab from as they pleased
  - Sharing a book with a group of friends (first-come-first-serve)
- Whenever these questions arose, the discussion was lively
  - Suggestions, work-arounds, kludges, "me too"s
  - A shared bookshelf (like BookBits) would have met these recurring needs
- There was Buzz around this issue! But we ignored it. Or didn't find it.
  - Labor intensive to continually keep up with forum activity

WVCorp has the information to help address this need, but not enough resources (labor) to use the information effectively.

In a real presentation, use a screenshot of a relevant forum discussion.

2- Stating the project goal

Goal: Catch it Early

- Predict which topics on our product forums will have persistent buzz
  - Features customers want
  - Existing features users have trouble with
- Persistent buzz, not ephemeral or trendy issues
  - Persistence = real, ongoing customer need

State the project goal in terms of the business motivation.

In a real presentation, use a screenshot of a relevant forum discussion.

Stating the project's results

3- Describing the project and its results

Pilot Study

- Collected three weeks of data from forum
- Trained model on Week 1 to identify which topics will buzz in Weeks 2/3
  - Buzz = Sustained increase of 500+ active discussions in topic/day, relative to Week 1, Day 1
  - Compared predicted results to topics that actually buzzed
  - Feedback from team of five product managers—how useful were the results?

Briefly describe how the project was run.

Results

- Reduced manual scan of forums by over a factor of 4
  - Scan 184 topics—not 791!
  - PMs: 75% of identified topics produced "valuable insight"
- Found 84% of about-to-buzz topics
- Low (20%) false positive rate

	Predicted No Buzz	Predicted Buzz	
No Buzz	579	35	614
Buzz	28	149	177
Total	607	184	791

topics predicted to buzz that didn't

about-to-buzz topics that were missed

topics the PMs can skip

topics the PMs have to review

Representative end users thought the model's output was useful.

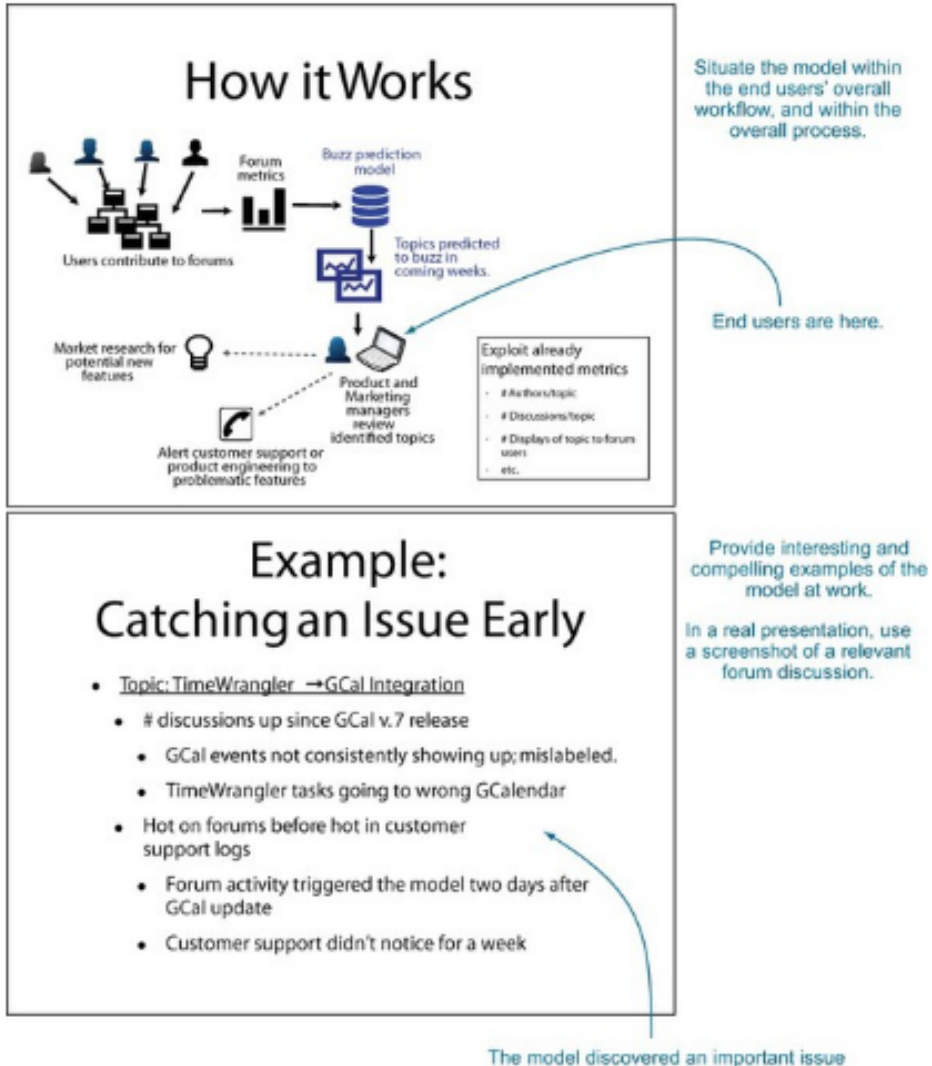
State the results up front.

State the results in terms of how they affect the end users (product managers).

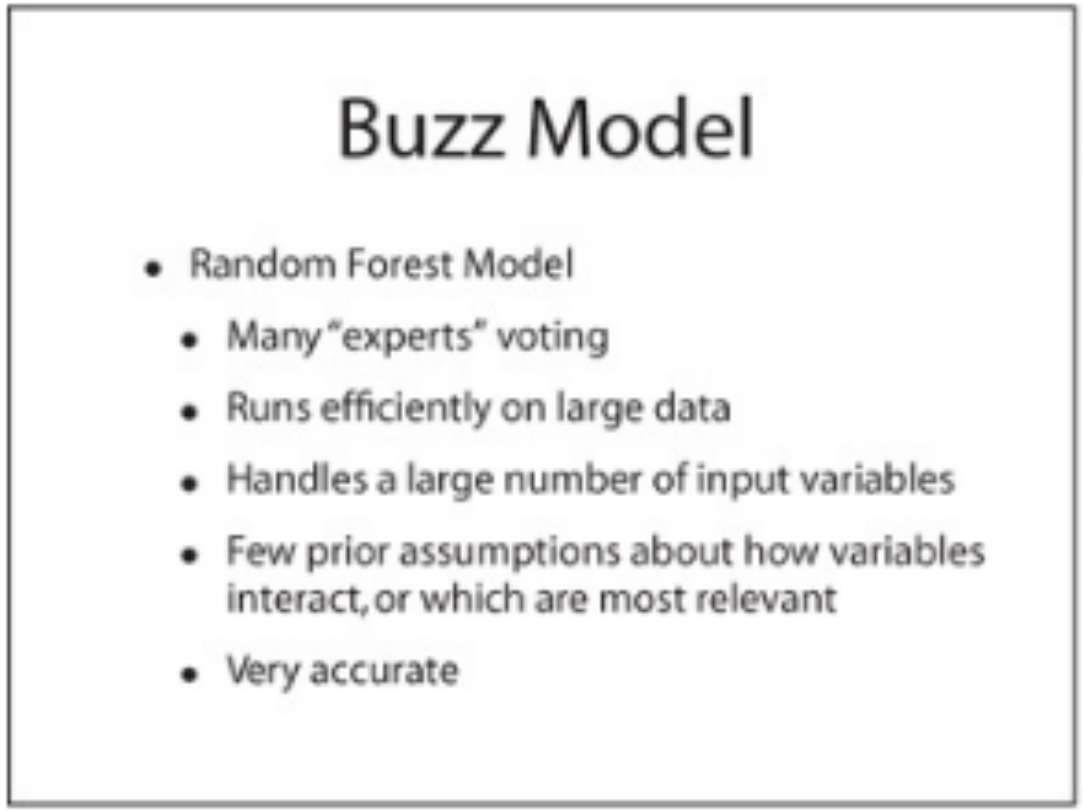
The model reduces the end users' workload by zeroing in on what they need to look at.

Filling in the details

4- Discussing your work in more detail



5- Optional slide on the modeling method



An optional slide briefly discusses details of the modeling method.

Making recommendations and discussing future work

6- Discussing future work

Project sponsor presentation takeaways

- Keep it short.
- Keep it focused on the business issues, not the technical ones.
- Your project sponsor might use your presentation to help sell the project or its results to the rest of the organization. Keep that in mind when presenting background and motivation.
- Introduce your results early in the presentation, rather than building up to them.

Presenting your model to end users

1. Summarize the motivation behind the project, and its goals.

2. Show how the model fits into the users' workflow (and how it improves that workflow).

3. Show how to use the model.

Let's explore each of these points in turn, starting with project goals.

1- Summarizing the project's goals

- The users already know that they want to find buzz; our model will help them search more effectively.

Motivation for project

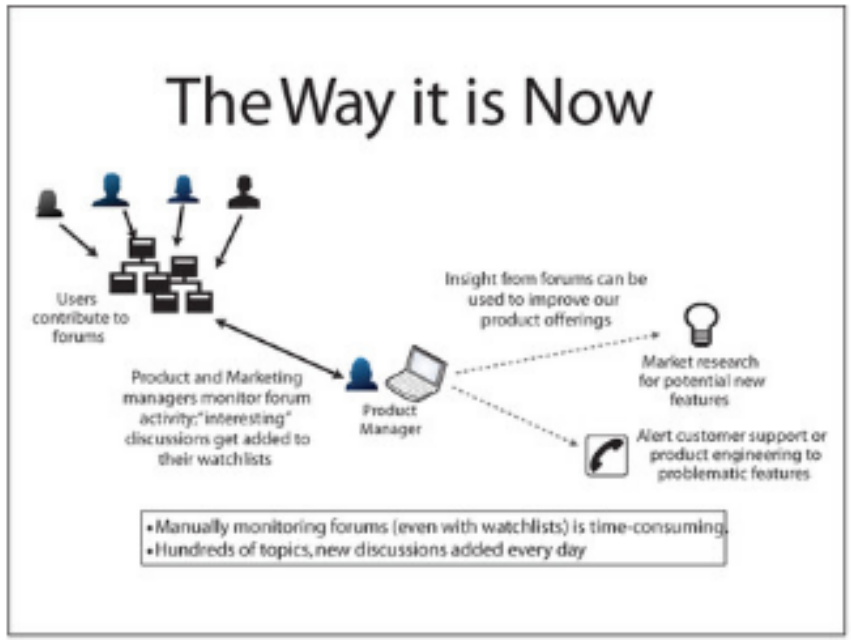
Provide motivation for the work from the end user's perspective: help them find useful buzz faster.

In a real presentation, you might use a screenshot of a relevant forum discussion.

Showing how the model fits the users' workflow

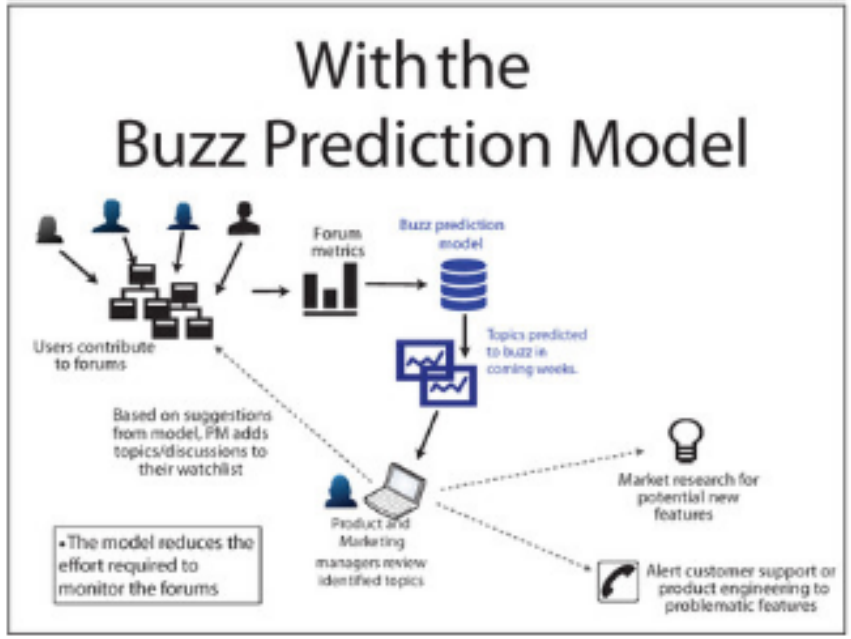
- A good way to do this is to give before-and-after scenarios of a typical user workflow.

User workflow before and after the model



Compare the end users' day-to-day work process before and after the introduction of your model.

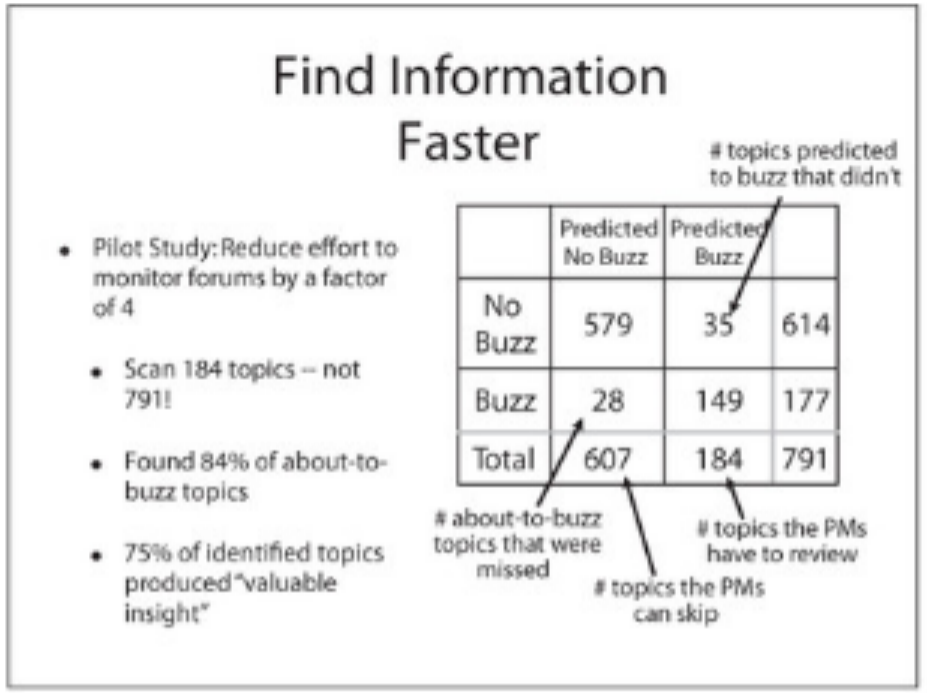
Before: time-consuming.



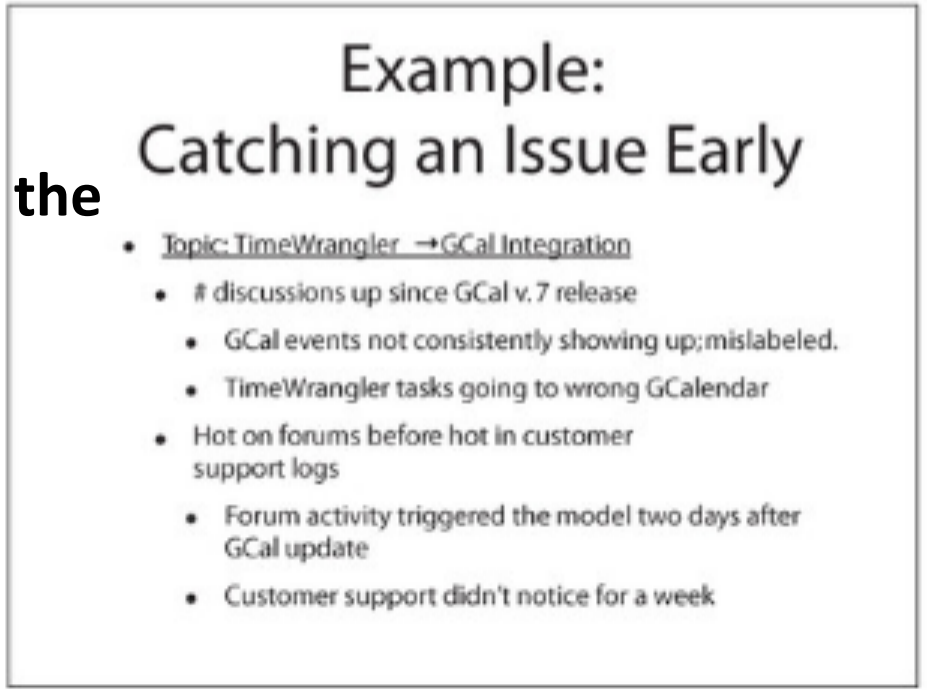
After: more focused, less time-consuming.

- The next slide (figure , top) uses the pilot study results to show that the model can reduce the effort it takes to monitor the forums, and does in fact provide useful information.

Present the model's benefits from the users' perspective.



State the results from the end user's perspective: manual effort is reduced, and the model's suggestions are correct and valuable.



Show interesting and compelling examples of the model at work.

In a real presentation, you might use a screenshot of a relevant forum discussion.

Provide technical details that are relevant to the users.

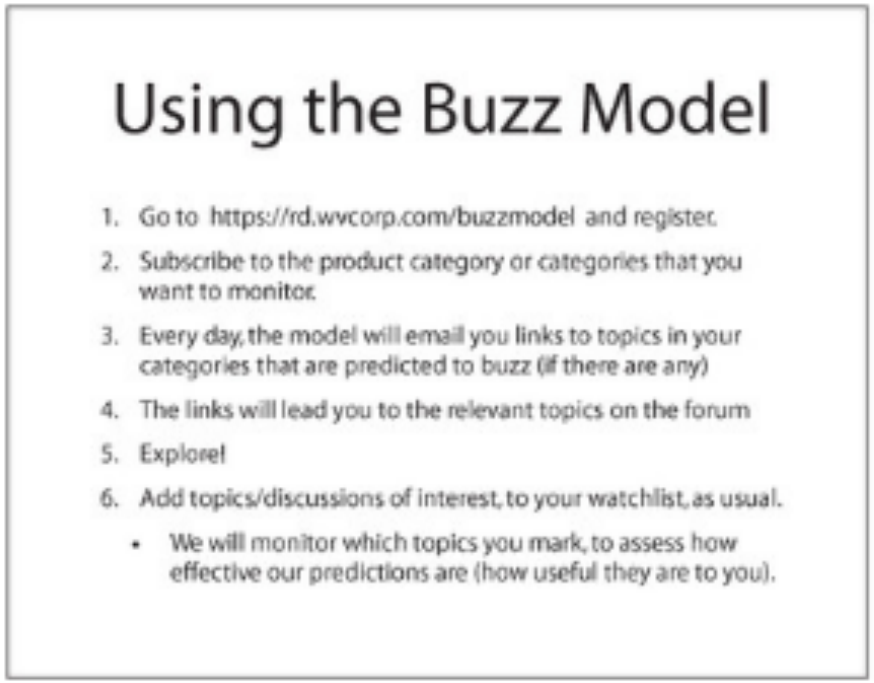
- You may also want to fill in more details about how the model operates.
- For example, users may want to know what the inputs to the model are (figure), so that they can compare those inputs with what they themselves consider when looking for interesting information on the forums manually.

The end users will likely be interested in the inputs to the model (to compare with their own mental processes when they look for buzz manually).

Showing how to use the model

- The slide in figure describes how a product manager will interact with the buzz model.
- In this example scenario, we're assuming that there's an existing mechanism for product managers to add topics and discussions from the forums to a watchlist, as well as a way for product managers to monitor that watchlist.
- The model will separately send the users notifications about impending buzz on topics they're interested in.

Describe how the users will interact with the model.

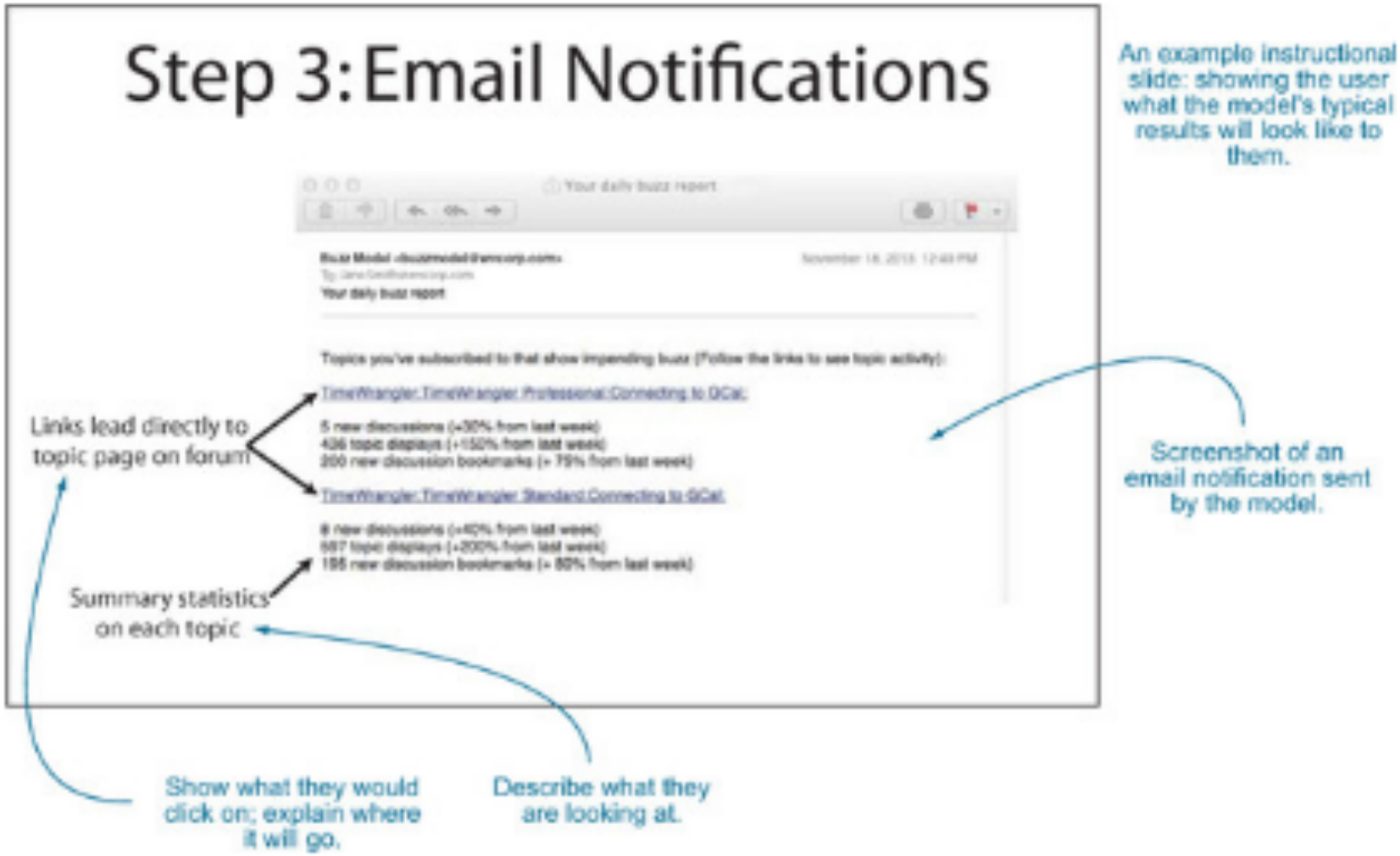


Show the users how to interact with the model.

In a real presentation, each point would be expanded with step-by-step detailed instructions and appropriate screenshots.

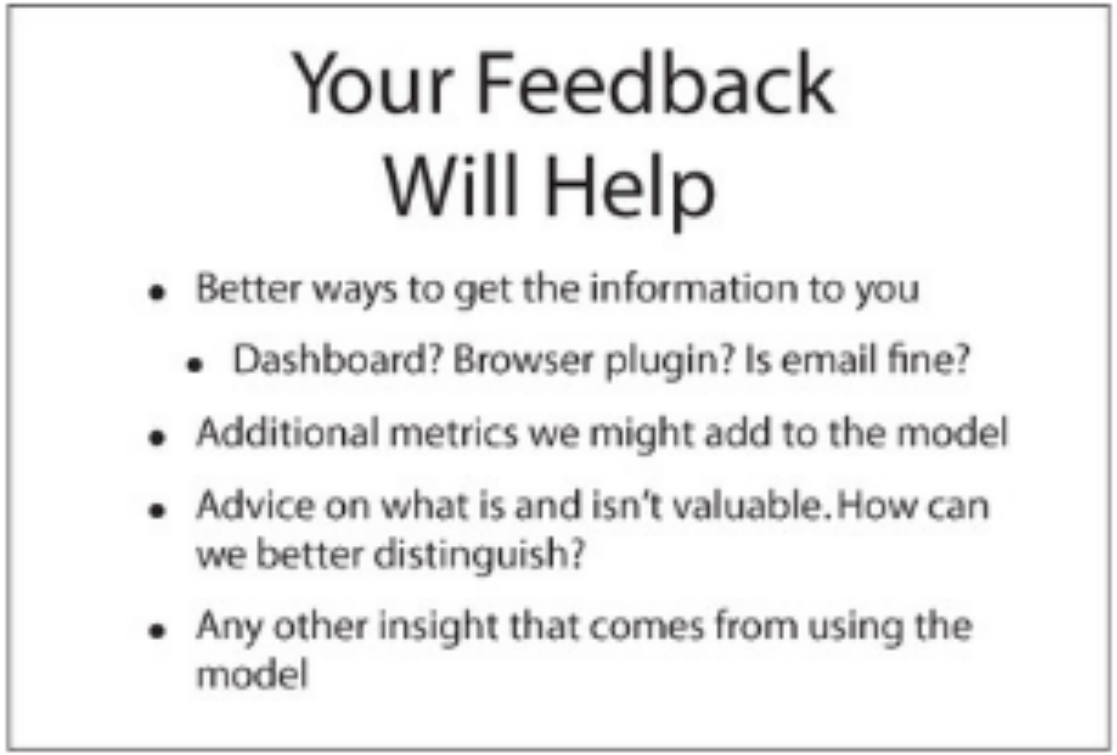
- In a real presentation, you'd then expand each point to walk the users through how they use the model: screenshots of the GUIs that they use to interact with the model, and screenshots of model output.
- We give one example slide in figure a screenshot of a notification email, annotated to explain the view to the user.

An example instructional slide



- included a slide that asks the users for feedback on the model, once they've been using it in earnest. This is shown in figure . Feedback from the users can help you (and other teams that help to support the model once it's operational) to improve the experience for the users, making it more likely that the model will be accepted and widely adopted.

Ask the users for feedback.



Enlist the end users' help in improving the model (and the overall workflow): ask for feedback.

End user presentation takeaways

- Your primary goal is to convince the users that they want to use your model.
- Focus on how the model affects (improves) the end users' day-to-day processes.
- Describe how to use the model and how to interpret or use the model's outputs.

Presenting your work to other data scientists

- Presenting to other data scientists gives them a chance to evaluate your work and gives you a chance to benefit from their insight.
- They may see something in the problem that you missed, and can suggest good variations to your approach or alternative approaches that you didn't think of.

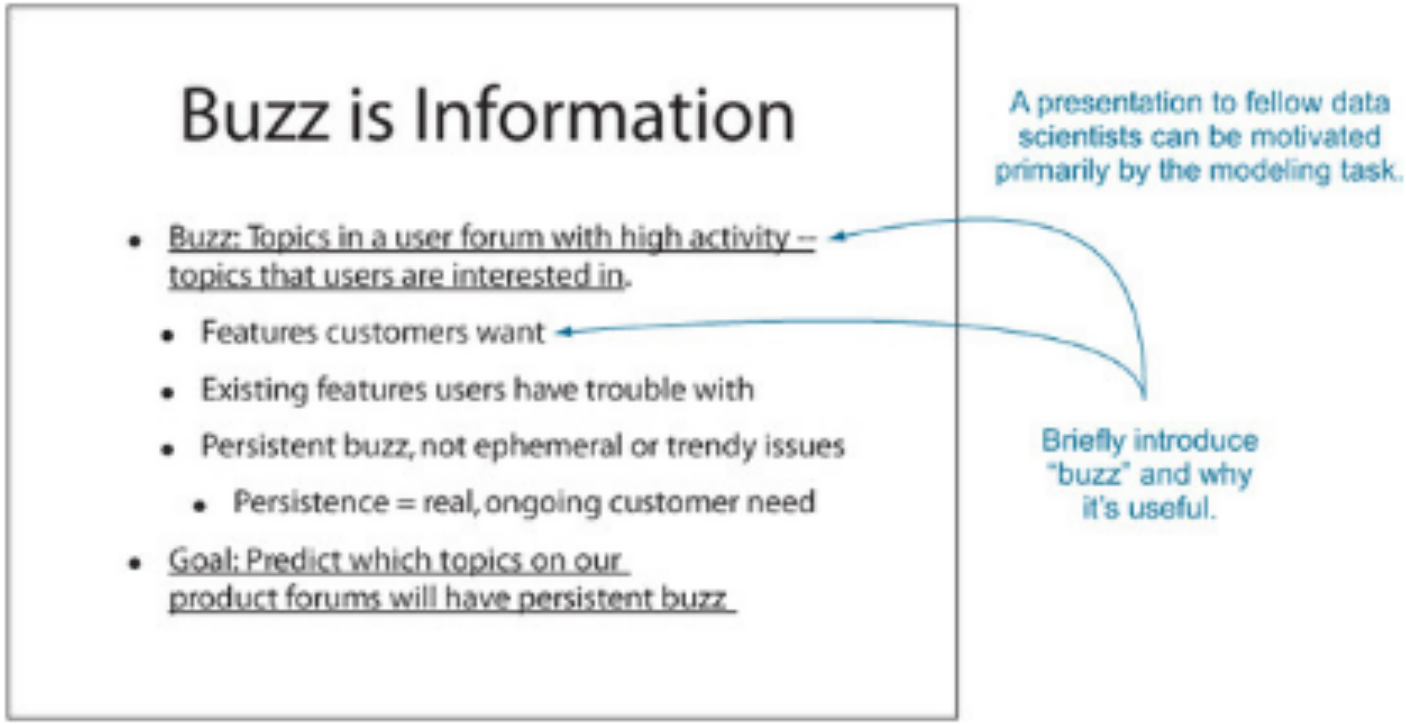
presentation to your peers generally has the following structure:

1. Introduce the problem.
2. Discuss related work.
3. Discuss your approach.
4. Give results and findings.
5. Discuss future work.

1- Introducing the problem

- This approach is best when you're presenting to other data scientists within your own organization, since you all share the context of the organization's needs.
- When you're presenting to peer groups outside your organization, you may want to lead with the business problem

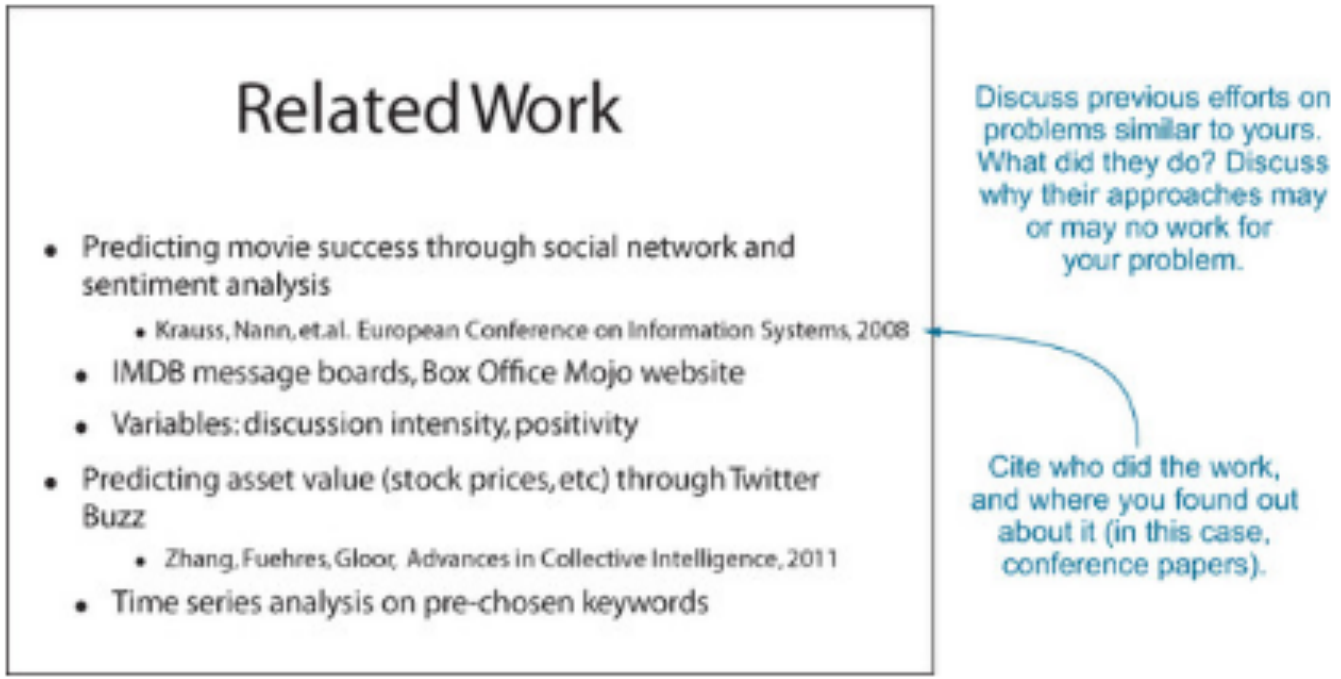
Introducing the project



2- Discussing related work

a related work slide is an opportunity to discuss other approaches that you considered, and why they may not be completely appropriate for your specific problem.

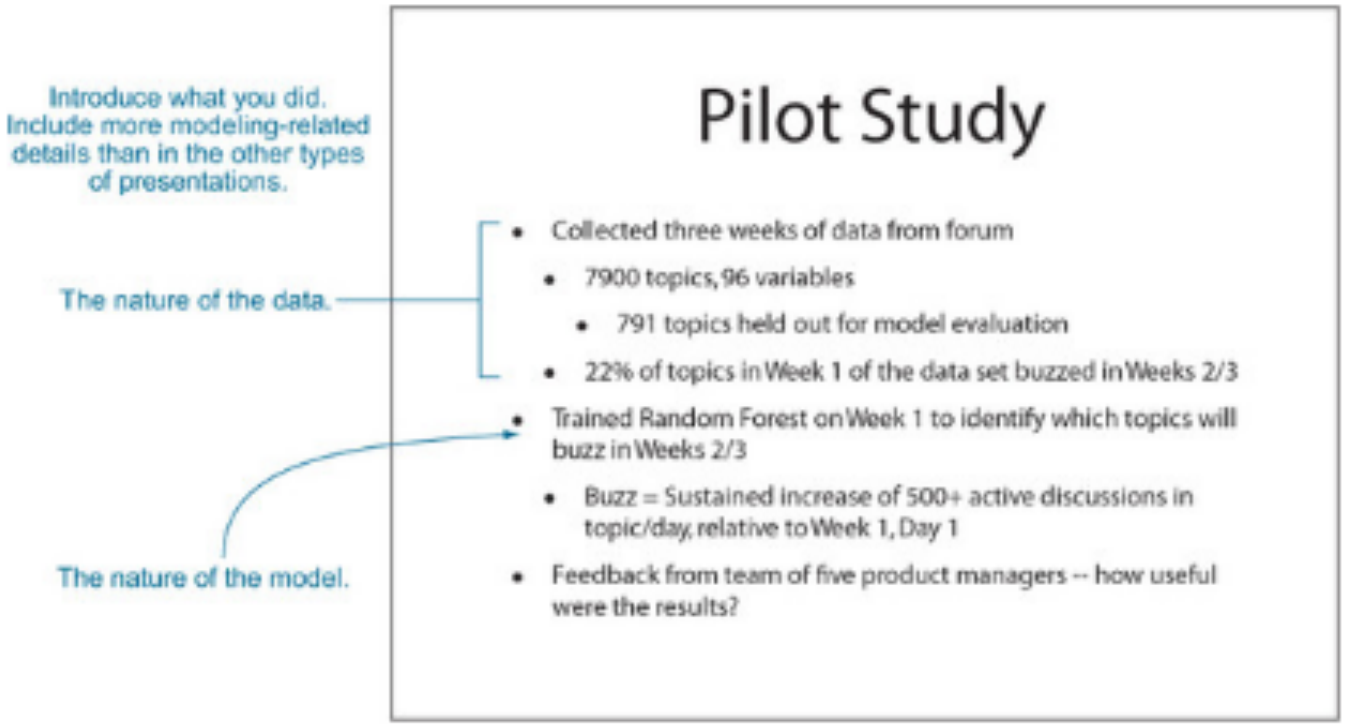
Discussing related work



3-Discussing your approach

introduces the pilot study that we conducted, the data that we used, and the modeling approach we chose. It also mentions that a group of end users (five product managers) participated in the project; this establishes that we made sure that the model's outputs are useful and relevant.

Introducing the pilot study



3-Discussing your approach

In this scenario, the dataset didn't have the right variables—it would have been better to do more of a time-series analysis, if we had the appropriate data, but we wanted to start with metrics that were already implemented in the product forums' system. Be up-front about this.

Discussing model inputs and modeling approach

Model Variables

- We started with metrics already monitored by system.
  - #Authors/topic
  - #Discussions/topic
  - #Displays of topic to forum users
  - Average #contributors to a discussion in the topic
  - Average discussion length in a topic
  - How often a discussion in a topic is forwarded to social media
- Obviously problematic -- only point measurements
  - Ideally, we want to measure evolution
    - Are, e.g. the number of authors increasing/ decreasing? How fast?
  - Time-series analysis
  - How well can we do with what we have?

Introduce the input variables (and issues with them).

Random Forest Model

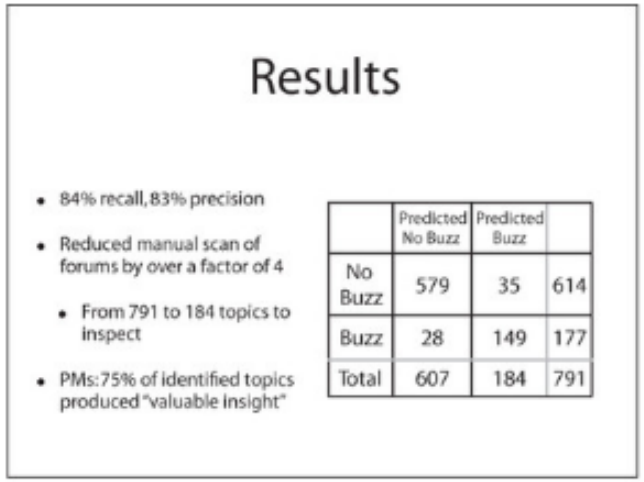
- Efficient on large data, large number of input variables
- Few prior assumptions on variable distribution/ interactions
- We limited complexity to reduce overfit
  - 100 nodes/tree maximum
  - Minimum node size 20
  - More data would eliminate the need for these steps

Introduce the model, why you chose it, and issues with it.

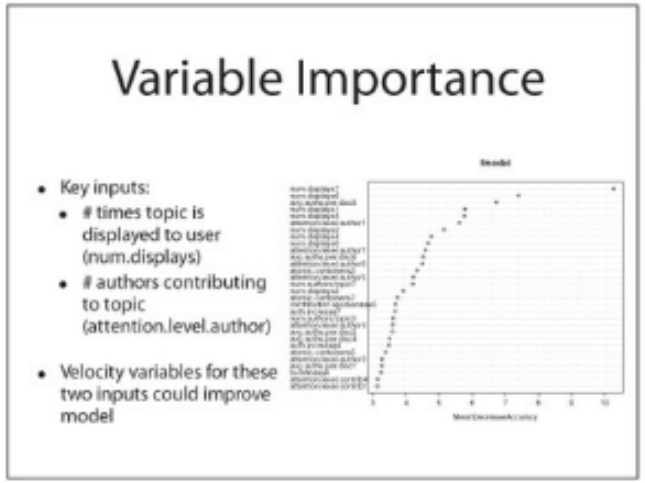
Discussing results and future work

- Discuss our model's performance (precision/ recall) and also confirm that representative end users did find the model's output useful to their jobs.
- The bottom slide of figure shows which variables are most influential in the model (recall that the variable importance calculation is one side effect of building random forests). In this case, the most important variables are the number of times the topic is displayed on various days and how many authors are contributing to the topic.

Showing model performance



Show your results: model performance and other outcomes.

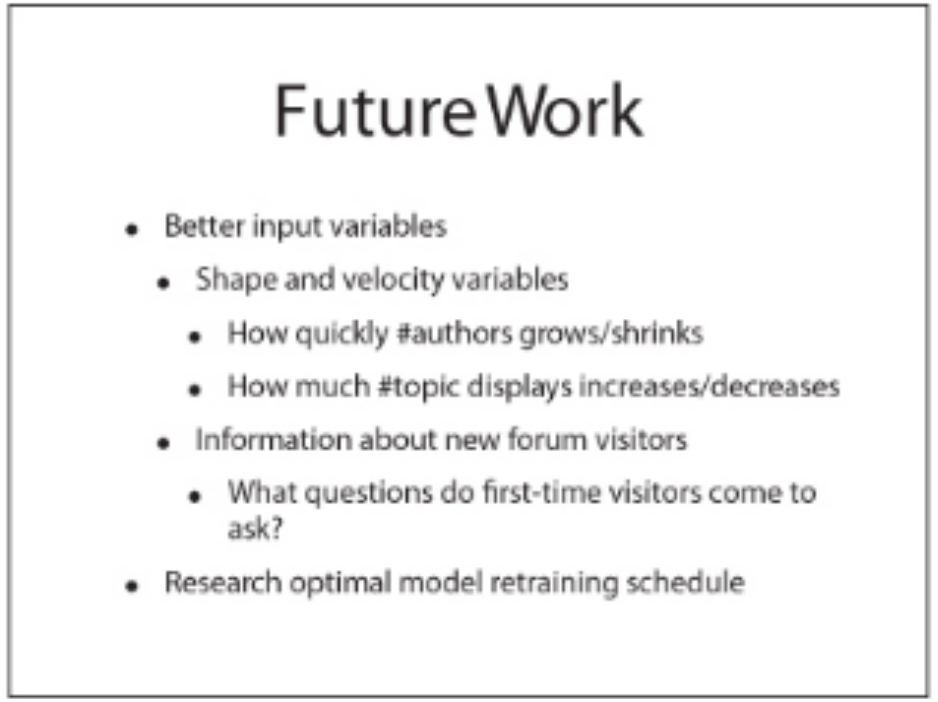


Discuss other key findings, like which variables were most influential on the model.

Discussing results and future work

- End the talk with a discussion of possible improvements and future work.
- Some of the points on the future work slide—in particular the need for velocity variables—come up naturally from the previous discussion of the work and findings. Others, like future work on model retraining schedules, aren't foreshadowed as strongly by the earlier part of the talk, but might occur to people in your audience and are worth elaborating on briefly here.

Discussing future work



Discuss future work.

Peer presentation takeaways

- A peer presentation can be motivated primarily by the modeling task.
- Unlike the previous presentations, the peer presentation can (and should) be rich in technical details.
- Be up-front about limitations of the model and assumptions made while building it. Your audience can probably spot many of the limitations already.

Key takeaways

- Presentations should be organized and written with a specific audience and purpose in mind.
- Organize your presentations to declare a shared goal and show how you're meeting that goal.
- Some presentations are more technical than others, but all should be honest and share convincing work and interesting results.