Testing to Improve User Response of Crowdsourced S&T Forecasting System

Sponsors:

Charles Twardy, GMU C4i Center, Adam Siegel, Inkling Markets



Team members: Kevin Connor Andrew Kreeger Neil Wood



Project Background

SciCast central premise: "the collective wisdom of an informed and diverse group is often more accurate at forecasting the outcome of events than that of one individual expert."





^{*}SciCast is run by George Mason University and sponsored by the U.S. Government.

SciCast[‡] is a research project that forecasts outcomes of key issues in science and technology.

SciCast Overview

What will be the total value of U.S. cherry production for the 2013 marketing season, in thousands of USD?



Via SciCast, users can make and change their forecasts at any time on a published question.

SciCast Overview

SciCast functions like a real-time indicator of what our participants think is going to happen.



Forecasts made by SciCast Users are aggregated to provide predictions on questions.

Project Goals

- Problem Statement:
 - In general, crowdsourcing sites require a large and diverse number of participants making forecasts.
 - SciCast is no exception and our project sponsors would like to see more forecasts being made on SciCast.
- Propose and Evaluate Web UI (User Interface) design modifications to:
 - 1. Increase the user participation rate
 - i.e., increase average number of forecasts made by each user
 - 2. Increase the size of the SciCast user base
 - i.e., increase SciCast registration rate
- Proposed Web UI design modifications
 - 1. Recommender Box -- Used to increase SciCast user participation rate
 - 2. Updated Splash Page -- Used to increase SciCast user registration rate

Experimental Approach

- Design and run a hypothesis test on the recommender box
 - Use test to determine if there is an increase in user participation
- Design and conduct a focus group study
 - Use study to discover problems areas with the SciCast site and to discover potential areas for improvement
- Design and run a hypothesis test on the splash page
 Use study to determine if there is an increase in user registration
- Hypothesis tests will employ A/B or A/B/C types of tests.

A/B Testing Overview

- A/B website testing refers to testing by comparing two (or more) website versions (an A version and a B version) where differences between the versions is minimal
- Users going to the website are assigned to one version of the website
- Differences in behavior of the users of the two sites can be attributed to the differences between the sites
- It is important to have a large random sample to ensure the testing data reflects the population

User Participation Rate



[‡]Data for the chart is extracted through the SciCast Datamart Interface.

SciCast Challenge

- Average user participation is low (i.e. most users make less than 5 forecasts)
- Most of the forecasts are from roughly 5% of the user base
- A low participation rate causes a reduction in the total number of forecasts and results in a lack of diversity in forecasts
- Higher participation from a more diverse group could improve the accuracy of SciCast forecasts

Increasing user participation rate could improve the accuracy of SciCast forecasts

Recommender Box Description



YOUR RECOMMENDED QUESTIONS		
V d t	What will be the total U.S. almond utilized production during the 2013 marketing season, in thousands of cons?	
	eading possible answer Between 1780 and 1940 thousand tons	Ø
C	CURRENT FORECAST	
2	24%	
U	inchanged today	
	WHAT DO YOU THINK?	
Pacammandar Bay		

- Recommender Box
- The recommender box contains a list of questions considered relevant to the user
- List is created by an algorithm developed by the SciCast team

Our team was asked to evaluate the impact of a recommender box on user participation.

Recommender Box Experimental Design

- Experimental Goal, answer the following questions:
 - 1. Does the recommender box increase the number of user forecasts?
 - 2. Does the algorithm that creates recommendations work?
 - 3. Why or why not?
- Experimental Technique:
 - A/B/C Hypothesis Test
 - Used to answer questions 1 and 2
 - Quantitative analysis method
 - Focus Group Test
 - Used to answer question 3
 - Qualitative analysis method

Our project team designed a quantitative and a qualitative test to evaluate the impact of the recommender box on user participation.

A/B/C Hypothesis Test

- Each SciCast User will be directed to one of three experimental groups:
 - A. Control Group: No changes with respect to the current site
 - B. Treatment Group: Recommender box providing recommended questions
 - C. Treatment Group: Recommender box providing random questions
- Users will be assigned to A,B and C groups using stratified sampling
- Will use hypothesis testing to determine if there are differences between the groups
 - Currently planning to use the Student's T test.
 - May switch to rank-sum or Kolmogorov-Smirnoff tests if the distributions do not meet the parametric assumptions for a normal distribution.



Experimental Metrics

- Metrics will be measured by using Google Analytics with the SciCast website
- Preliminary list of metrics:
 - Number of times a user clicked a question in the recommender
 - Number of times a user provided a forecast on a question reached through the recommender box
 - Number of times a user provided a forecast for a question reached external to the recommender box
 - Recommender's ranking of questions selected via the recommender box
 - Recommender's ranking of questions selected external to the recommender box
- Additional metrics may be added per sponsor direction

Experiment Status and Future Work

- Recommender box experiment has been designed and approved by project sponsor
- Recommender box experiment put on hold until the recommender can be fully integrated into the SciCast production site
- Future work (for a future class project)
 - Implement and run recommender box experiment in Google Analytics

Focus Group Background

- Sponsor's requested focus group to supplement A/B testing
- Focus group could answer:
 - Why A/B testing succeeded or failed
 - Why users are or are not drawn to the SciCast site
- Testing involving humans subjects required HSRB approval
 - HSRB approval required experimental design application and HSRB training

Focus Group Experiment

- Purpose and goal of SciCast site was explained to volunteers
- Volunteers then:
 - Created accounts on the test site
 - Explored site
 - Found question of interest
 - Made a prediction
 - Answered questionnaire about their experience
- These activities were timed with the goal of finding activities that the volunteers struggled with

Focus Group Results

- Users seemed confused about the purpose of the SciCast site
- Users had difficulty finding questions that interested them, or they felt they could answer
 - Trouble finding questions implies that a recommender box would improve participation
- Users had little trouble creating an account, navigating through the site or making a prediction
- Users failed to notice the recommender box

A recommender box will improve the site, but work may be required on drawing attention to recommended questions

Splash Page Background

- Due to delay in recommender box testing, team shifted focus to splash page testing
- Sponsor wanted to know if adding <u>sample questions</u> to the splash page would have an effect on user behavior
- Performed power analysis to determine expected experiment length
- Utilized Google Analytics to perform A/B testing
- Measured bounce rate to determine if splash page changes had an effect



Splash Page Results

- Experiment ran for 15 days with 2,576 total sessions for the experiment
- Due to the Multi-Armed Bandit approach for splitting site traffic, the original splash page had 719 sessions with a bounce rate of 4.03%
- The new splash page, Variation 1, had 1,857 sessions with a bounce rate of 3.02%



Splash Page Conclusions

- Based on the A/B testing results, we concluded that the proposed splash page caused a 25% reduction in the bounce rate
- Google Analytics ended the experiment without declaring a "winner" but...
- 90.9% confident the new splash page will lower bounce rates



Adding sample questions to the splash page increases the user interaction rate

Final Conclusions

- Team learned a lot along the way
- Successfully designed a future A/B/C test for the recommender box
- Successfully designed and carried out a focus group study
- Successfully designed and carried out an A/B test on the splash page
- Recommend:
 - SciCast implement recommender box
 - SciCast implement new splash page
 - SciCast continue to utilize A/B testing

Thank You & Questions?