10 Academy Batch 3

Weekly Challenge: Week 4

This case will be used for the week 4 of training for Batch 3.

A/B Hypothesis Testing: Ad campaign performance

Business objective

An advertising company is running an online ad for a client with the intention of increasing brand awareness. The advertiser company earns money by charging the client based on user engagements with the ad it designed and serves via different platforms. To increase its market competitiveness, the advertising company provides a further service that quantifies the increase in brand awareness as a result of the ads it shows to online users. The main objective of this project is to test if the ads that the advertising company runs resulted in a significant lift in brand awareness.

Project Overview

SmartAd is a mobile first advertiser agency. It designs Intuitive touch-enabled advertising. It provides brands with an automated advertising experience via machine learning and creative excellence. Their company is based on the principle of voluntary participation which is proven to increase brand engagement and memorability 10 x more than static alternatives.

SmartAd provides an additional service called Brand Impact Optimiser (BIO), a lightweight questionnaire, served with every campaign to determine the impact of the creative, the ad they design, on various upper funnel metrics, including memorability and brand sentiment.

As a data scientist in SmartAd, one of your tasks is to design a reliable hypothesis testing algorithm for the BIO service and to determine whether a recent advertising campaign resulted in a significant lift in brand awareness.

Data

The BIO data for this project is a "Yes" and "No" response of online users to the following question

Q: Do you know the brand SmartAd?

O Yes
O No

This is a test run and the main objective is to validate the hypothesis algorithm you built. SmartAd ranthis campaign from 3-10 July 2020. The users that were presented with the questionnaire above were chosen according to the following rule:

Control: users who have been shown a dummy ad
Exposed: users who have been shown a creative, an online interactive ad, with the SmartAd brand.

The data is available for download here.

The data collected for this challenge has the following columns

- **auction_id:** the unique id of the online user who has been presented the BIO. In standard terminologies this is called an impression id. The user may see the BIO questionnaire but choose not to respond. In that case both the yes and no columns are zero.
- **experiment**: which group the user belongs to - control or exposed.
- **date**: the date in YYYY-MM-DD format
- **hour**: the hour of the day in HH format.
- **device_make**: the name of the type of device the user has e.g. Samsung
- **platform_os:** the id of the OS the user has.
- **browser**: the name of the browser the user uses to see the BIO questionnaire.
- **yes**: 1 if the user chooses the "Yes" radio button for the BIO questionnaire.
- **no**: 1 if the user chooses the "No" radio button for the BIO questionnaire.

Learning Outcomes

Skills:
- Statistical Modelling
- Using core data science python libraries pandas, matplotlib, seaborn, scikit-learn
- Linear regression
- Decision Trees
- XGBoost

Knowledge:
- Data exploration
- Hypothesis testing
- Machine learning
- Hyperparameter tuning
- Model comparison & selection
- experiment analysis

Communication:
- Reporting on statistically complex issues

Team

Instructor: Yabebal Fantaye with Jean-Henock
Tutors: Sebastian, Abla, Usman, Moustapha

Key Dates

- Discussion on the case - 1130 Rwanda time on Monday 10 August 2020. Use #all-week4 to pre-ask questions.
- Interim Solution - 2000 Rwanda time on Tuesday 11 August 2020.
- Final Submission - 2000 Rwanda time on Saturday 15 August 2020

Grading for the week

There are 100 points available for the week.

20 points - community growth and peer support. This includes supporting other learners by answering questions (Slack), asking good questions (Slack), participating (not only attending) daily standups (GMeet) and sharing links and other learning resources with other learners.

25 points - presentation and reporting.

5 points - interim submission

20 points for the final submission. This is measured through:

- Clarity of writing (10 points)

- Clarity of structure and message including appropriate usage of graphs (5 points)
- Professionalism/production value (free of spelling errors, use of same font, well produced) (5 points)
- Balance between being 'full of information' and 'easy to understand' (5 points)

55 points

10 points - Interim submission

45 points -Final submission

Validity of model and recommendations (25 points)

Quality of code (20 points)

The following describes, at a high level, elements required for full marks.

Big Picture

- The objective of the work is clearly stated
- The stated objective is correct (the reporter has understood the task)

Details

- Consistent Voice in a section (1st person or 3rd person at least)
- The work is clearly motivated (e.g. main challenge this work directly or indirectly addresses))
- Data source clearly described
- Data structure clearly outlined (e.g. date ranges stated, condition at which data collected, etc.)
- Method clearly outlined
- Challenges encountered and addressed stated
- Evaluation metric clearly outlined (if appropriate)
- Valid insights drawn
- Valid conclusions drawn

Notebook (code) Structure

The main python authority for code style guide is the PEP 8 guideline. All coding styles are guided accordingly. We also use PEP 20 - The Zen Of Python - to make all other coding based judgments. Some important highlights are as follows.

Markdown

- Have section headings
- Have a basic explanation of what the code in each section does
- Have a good explanation of the approach taken for each section of code
- All sections and subsections have headers and have a reasonable explanation

Variable Naming

- Consistent. Function names should be lowercase, with words separated by underscores as necessary to improve readability.
- Variable names follow the same convention as function names.
- Standard (is followed by comment OR descriptive OR creative OR reasonable)

Python Functions

- Have at least one function
- Have at least one class
- Have more than one functions

Figures

Our guideline here is the infamous Google Material Design guideline which is the global definition for UI/UX and Data Visualisations.

Axes

- Appropriate font size (readable)
- Readable titles
- Has units
- Has legend (explains multiple linestyle, colors, markers are used)
- Has caption (explains what the plot is about)

Type of Figure

- Appropriate for the data
- Innovative

Github

- Has a basic readme that explains what the repository is about
- Has a detailed description
- Has frequent commits
- Is an installable package
- Has requirements.txt or similar to reproduce work

Style

- Uniform across pages and slides
- Pleasing density (low better), font, color and format. (for slides this guideline)

Community

- Supporting other learners by answering questions
- Asking good questions
- Participating (not only attending) daily standups
- Sharing links and other resources with other learners

Badges

Each week, one user will be awarded one of the badges below for the best performance in the category below.

In addition to being the badge holder for that badge, each badge winner will get +20 points to the overall score.

Visualization - quality of visualizations, understandability, skimmability, choice of visualization
Quality of code - reliability, maintainability, efficiency, commenting - in future this will be CICD
Innovative approach to analysis -using latest algorithms, adding in research paper content and other innovative approaches
Writing and presentation - clarity of written outputs, clarity of slides, overall production value
Most supportive in the community - helping others, adding links, tutoring those struggling

The goal of this approach is to support and reward expertise in different parts of the Data Scientist toolbox.

Late Policy

Our goal is to prepare successful learners for the work and submitting late, when given enough notice, shouldn't be necessary.

For interim submissions, those submitted 1-6 hours late will receive a maximum of 50% of the total possible grade. Those submitted >6 hours late may receive feedback, but will not receive a grade.

For final submissions, those submitted 1-24 hours late, will receive a maximum of 50% of the total possible grade. Those submitted >24 hours late may receive feedback, but will not receive a grade.

When calculating the leaderboard score:
- From week 4 onwards, your lowest week's score will not be considered.
- From week 8 onwards, your two lowest weeks' scores will not be considered.

Instructions

Objectives:

The global (business) objective is divided into 4 sub-objectives that overall guides the workflow

- Setting up A/B testing framework
- Validating the data validity
- Performing A/B testing with classical, sequential and Machine learning methods
- Extracting statistically valid insights in relation to the business objective

Why this project?

Hypothesis testing is the cornerstone of evidence based decision making. The A/B testing framework is the most used statistical framework for making gradual but important changes in every aspect of today's business. Please read A Refresher on A/B Testing to get a rich business and historical context.

Detailed instruction:

Here is the summary of tasks you will perform.

- Read this document carefully and make sure you have understood the business and data analysis objectives.
- Obtain the data from here
- Read the main reference paper and blog entries. We highly recommend you get a good understanding of the subtleties involved in the A/B testing framework. In particular why is it important to not perform the classical A/B testing analysis while the experiment is running? Study the recommended Kaggle kernels to get a better understanding.
- Understand the data. Make visualisation and ensure you understand how the data is collected and what each features are.
- Attempt all tasks defined below.
- Upload your jupyter notebook to your Github public repository.
- For the interim submission, only a PDF report is required & for final submission both link to your GitHub repository & your PDF report are requested .
- If you have any questions or confusions regarding what you are expected to do in this project or how to submit, please contact the team

Task 1: Framework

Task 1.1: Setting up an A/B testing strategy

Please answer the following questions as it relates to this case.

- Which online users belong to the control and exposed groups?
- How are the users targeted?
- Could we use the counts of yes and no answers to make a judgement on which experiment is performing better? For example if #yes > #no for the exposed group than the control group, could we declare that the ad had a significant impact Why or why not?
- What is the statistical process that generates the data? Which kind of statistical model will you use if you were to simulate the data?
- Assessment of the statistical significance of an A/B test is dependent on what kind of probability distribution the experimental data follows. Given your answer above, which statistical tests (z-test, t-test, etc.) are appropriate to use for this project?
- In classical (frequentist) A/B testing, we use p-values to measure the significance of the experimental feature (being exposed to an ad in our case) over the null hypothesis (the hypothesis that there is no difference in brand awareness between the exposed and control groups in the current case). How are p-values computed? What information do p-values provide? What are the type-I and type-II errors you may have in the analysis? Can you comment to which error types p-values are related?
- How does the classical A/B testing (using z-test, f-test, etc.) framework work?
- How does sequential A/B testing work?
- What are some of the advantages of sequential A/B testing?
- How is A/B testing done using machine learning? What is the core idea behind this approach? In other words, what part of the machine learning analysis provides the insight regarding the high or no significance of the experimental feature?
- What are the pros and cons of using Machine learning to perform A/B testing?

Task 2: Analysis

Task 2.1 : Classic and sequential A/B testing analysis

- Perform data exploration to count unique values of categorical variables, make histogram, relational, and other necessary plots to help understand the data. For each of the plots you produce, write a description of what the plot shows in markdown cells.
- Perform hypothesis testing: apply the classical p-value based algorithm and the sequential A/B testing algorithm for which a starter code is provided..

- Are the number of data points in the experiment enough to make a reasonable judgement or should the company run a longer experiment? Remember that running the experiment longer may be costly for many reasons, so you should always optimize the number of samples to make a statistically sound decision.
- What does your A/B testing analysis tell you? Is brand awareness increased for the exposed group?

Task 2.2: Machine Learning

- In max three statements, make a problem formulation for machine learning and specify the target variable
- Split the data into 70% training, 20% validation, and 10% test sets.
- Based on the reading material provided, apply machine learning to the training data. Train a machine learning model using 5-fold cross validation the following 3 different algorithms:
  - Logistic Regression
  - Decision Trees
  - XGBoost
- Define the appropriate loss function for the model using the validation data.
- Compute feature importance - what's driving the model? Which parameters are important predictors for the different ML models? What contributes to the goal of gaining more "Yes" results?
- Which data features are relevant to predicting the target variable?
- Explain what the difference is between using A/B testing to test a hypothesis vs using Machine learning to learn the viability of the same effect?
- Explain the purpose of training using k-fold cross validation instead of using the whole data to train the ML models?
- What information do you gain using the Machine Learning approach that you couldn't obtain using A/B testing?

Task 2.3 : Reporting

- Prepare a presentation (20 slides max) to present your analysis to your company. This should include:
  - Objective of the study
  - Methods
  - Data
  - Results using both methods
  - Comparison of the two methods
  - Overall results
  - Recommendation and outcomes
  - Limitations of the analysis
  - References.

Interim Submission (Due Tuesday 11 August 2020 20hr Rwanda time)

- Share a report that addresses the points from task 1 (answer all questions in task 1.1.). Maximum of 3 pages - PDF format please. Prepare this in a format that you could share this as a learning exercise with 3rd-year students at your university.

Feedback

You may not receive detailed comments on your interim submission, but will receive a grade.

Final Submission (Due Saturday 15 August 2020 20hr Rwanda time)

- Link to your code in GitHub
- A presentation (20 slides max) covering all insights you gained from task 1 & task 2.

Feedback

You will receive comments/feedback in addition to a grade.

References

Key Papers and Blogs

- Classical A/B testing
    - http://sl8r000.github.io/ab_testing_statistics/
    - http://www.qubit.com/wp-content/uploads/2017/12/qubit-research-ab-test-results-are-illusory.pdf
    - https://projector-video-pdf-converter.datacamp.com/6165/chapter3.pdf
- Sequential testing
    - https://www.austinrochford.com/posts/2014-01-01-intro-to-sequential-testing.html
    - https://www.jstor.org/stable/2346379?seq=1
    - https://blog.rankdynamics.com/2015/10/27/the-proof-is-in-the-pudding/
- Machine Learning based A/B testing
    - A/B Testing with Machine Learning - A Step-by-Step Tutorial
- Python package
    - https://github.com/shansfolder/AB-Test-Early-Stopping
    - https://github.com/Testispuncher/Sequential-Probability-Ratio-Test
- Sequential testing R package
    - https://github.com/mdcramer/SPRT

Must Read

- Statistical Significance in A/B Testing – a Complete Guide
- A/B test with Python
- A Refresher on A/B Testing
- A/B Testing: Analysis of Credit Card Marketing Campaign | by Kailash Hari | Analytics Vidhya
- A/B Testing Statistics: An Easy-to-Understand Guide
- Sequential A/B Testing: Workflow and Advantages over Classic Experiments

Examples

- (Bio)statistics in R: Part #3
- Unit 3 - Hypothesis Testing
- Learning About User Retention - Meta Kaggle