A/B TESTING FOR INDUSTRIAL COMPANIES

A PROJECT REPORT

Submitted by

ANJALI AGRAWAL [Reg No: RA1811030010096]

Under the guidance of

Dr. S. Prabakeran

(Assistant Professor, Department of Networking And Communications)

in partial fulfillment of the requirements for the degree of

BACHELOR OF TECHNOLOGY

in

COMPUTER SCIENCE ENGINEERING

with specialization in CYBER SECURITY

DEPARTMENT OF NETWORKING AND COMMUNICATIONS COLLEGE OF ENGINEERING AND TECHNOLOGY
SRM INSTITUTE OF SCIENCE ANDTECHNOLOGY
KATTANKULATHUR- 603 203

MAY 2022

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

KATTANKULATHUR – 603 203

BONAFIDE CERTIFICATE

Certified that the 18CSP110L Internship report titled "**A/B TESTING FOR INDUSTRIAL COMPANIES**" is the bonafide work of **ANJALI AGRAWAL [Reg No: RA1811030010096]**, who carried out the project work under my supervision. Certified further, that to the best of my knowledge the work reported herein does not form any other project report or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE

SUPERVISOR
Dr. S. Prabakeran
Assistant Professor
Department of Networking and Communications

SIGNATURE

HEAD OF THE DEPARTMENT
Dr. Annapurani Panaiyappan.K
Professor
Department of Networking and Communications

Signature of the Internal Examiner

Signature of the External Examiner

Department of Computer Science and Engineering
SRM Institute of Science &Technology
Own Work* Declaration Form

This sheet must be filled in (each box ticked to show that the condition has been met). It must besigned and dated along with your student registration number and included with all assignments you submit – work will not be marked unless this is done.

To be completed by the student for all assessments

Degree/ Course : **B Tech. in CSE with spec. in Cyber Security**

Student Name : **Anjali Agrawal**

Registration Number : **RA1811030010096**

Title of Work : **A/B TESTING FOR INDUSTRIAL COMPANIES**

I / We hereby certify that this assessment compiles with the University's Rules and Regulations relating to Academic misconduct and plagiarism**, as listed in the University Website, Regulations, and the Education Committee guidelines.

I / We confirm that all the work contained in this assessment is my / our own except where indicated, and that I / We have met the following conditions:

- Clearly references / listed all sources as appropriate
- Referenced and put in inverted commas all quoted text (from books, web, etc)
- Given the sources of all pictures, data etc. that are not my own
- Not made any use of the report(s) or essay(s) of any other student(s) either past or present
- Acknowledged in appropriate places any help that I have received from others (e.g.fellow students, technicians, statisticians, external sources)
- Compiled with any other plagiarism criteria specified in the Course handbook / University website

I understand that any false claim for this work will be penalised in accordance with theUniversity policies and regulations.

DECLARATION:
I am aware of and understand the University's policy on Academic misconduct and plagiarism and I certify that this assessment is my / our own work, except where indicated by referring, and that I have followed the good academic practices noted above.
RA1811030010096 (Anjali Agrawal) Date:

ACKNOWLEDGEMENT

I express my humble gratitude to **Dr. C. Muthamizhchelvan**, Vice-Chancellor, SRM Institute of Science and Technology, for the facilities extended for the project work and his continued support.

I extend my sincere thanks to Dean E&T, SRM Institute of Science and Technology, **Dr. T.V. Gopal**, for his invaluable support.

I wish to thank **Dr. Revathi Venkataraman**, Chairperson, School of Computing, SRM Institute of Science and Technology, for her support throughout the period of the project work.

I am extremely grateful to our head of the department**, Dr. Annapurani Panaiyappan.K,** Professor, Department of Networking and Communications, SRM Institute of Science and Technology, for her suggestions and encouragement at all the stages of the project work.

I would like to convey my thanks to Panel Head, **Dr. S. Prabakeran,** Assistant Professor, Department of Networking and Communications, SRM Institute of Science and Technology, and program coordinator **Dr.M.B. Mukesh Krishnan,** Associate Professor, Department of Networking and Communications, SRM Institute of Science and Technology, for their inputs during the project reviews and support.

We register our immeasurable thanks to our Faculty Advisor**, Dr. M. Uma,** Associate Professor, Department of Networking and Communications, SRM Institute of Science and Technology, for leading and helping us to complete our course.

Our inexpressible respect and thanks to my guide, **Dr. S. Prabakeran,** Assistant Professor, Department of Networking and Communications, SRM Institute of Science and Technology, for providing me an opportunity to pursue my project under his/her mentorship. He provided me the freedom and support to explore the research topics of my interest. His passion for solving real problems and making a difference in the world has always been inspiring.

I sincerely thank the staff and students of the Networking and Communications Department, SRM Institute of Science and Technology, for their help during my research. Finally, we would like to thank my parents, our family members, and our friends for their unconditional love, constant support, and encouragement.

Anjali Agrawal

ABSTRACT

User-intensive software, like mobile applications and websites, heavily depends on interactions with many users and an unknown population. With the internet connectivity on such software, the website allows evaluating ideas and innovations using continuous experiments like A/B tests, split tests. We aim to study the application of A/B testing in various industrial contexts. We will also present a brief study on different statistical tests for different assumptions and solve them using the Bayesian algorithm. Randomized algorithms are used to address various software engineering problems. This type of algorithm gives different results with every run for the same problem instance. Therefore, a statistical test is important to prove the conclusions derived from the data.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ACRONYMS AND ABBREVIATION

MNCs	Multi-National Companies
ROI	Return of Investment
URL	Uniform Resource Locator
UI	User Interface
MVT	Multivariate testing
PnC	Permutation and Combination
SBSE	Stir Bar Sorptive Extraction
SQL	Structured Query Language
RDBMS	Relational Database Systems
VBFA/VBA	Visual Basic for Applications
PPC	Pay Per Click
HOC	Healthcare Organizations
CX	Customer Experience
HIPPA	Health Insurance Portability and Accountability Act
ASO	App Store Optimization
CPI	Consumer Price Index
LTV	Lifetime Value
CTR	Click-Through Rate
CTP	Click-Through Probability
VWO	Visual Website Optimizer
MDE	Minimum Detectable Effect

CHAPTER 1

INTRODUCTION

1.1 GENERAL

This paper focus on how A/B testing helps the industry understand their customer and increase customer engagement. This also helps them increase revenue generation and customer retention. Not every marketing campaign will provide favorable results, no matter how much research you conduct beforehand. As a result, A/B testing is an excellent approach for determining the most effective online advertising and marketing strategies for your company. It may be used to test anything, from website copy to sales letters and everything in between. Prior to spending your entire money on marketing materials that don't work, you can use this method to determine which version of your campaign performs best. While A/B testing might be time-consuming, the benefits outweigh the effort spent in the process. All things considered; well-executed A/B tests can make a significant impact on the effectiveness of your marketing campaigns. A campaign's return on investment can be increased by narrowing down and merging the most effective aspects of the promotion. This results in a lesser risk of failure, and most importantly, a more effective marketing plan. A/B test (controlled or randomized experiment) is the standard method to make data-driven decisions. It is the process of testing two or more versions of marketing assets, to identify which one performs better. It takes the output from digital marketing through the collection of data. By analyzing the result of each version, we get an understanding of what works and what doesn’t. It increases the chance of customer conversion and user engagement with the product.

1.2MOTIVATION FOR THE PROJECT

While working as an intern at Tiger Analytics, I was introduced to the A/B testing. With time I got to know about its use, benefits, and its importance for MNCs. I got an opportunity to work for a tech security service MNC. It was my first practical exposure to how it changes the company’s marketing strategy and decision-making system. By analyzing how users respond to a new change to what needs to be done to attract more users. We can make

decisions wisely by analyzing them using A/B tests. It is a real-time test and can be adjusted according to the manager's request. We can pre-calculate the time needed to perform the test and check its success rate and can also change the time duration of the test with low risk compared to other methods.

1.3 IMPORTANCE OF A/B TESTING

Accurate A/B tests can make a significant difference in the amount of money you make back. By conducting controlled testing and collecting empirical data, you can determine which marketing methods are the most effective for your organization and your product in specific situations.

In the event that one variation performs two, three, or even four times better than another without putting a significant amount of capital at risk, it is irresponsible to launch a promotion without first conducting extensive testing.

Testing, when done consistently, has the potential to significantly improve your results. In the long term, it is much easier to make judgments and develop more effective marketing plans if you understand what works and what doesn't (and have proof to support your conclusions).

Reduce bounce rates: The bounce rate of your website is one of the most crucial metrics to watch in order to evaluate its overall performance. There could be a variety of factors contributing to your website's high bounce rate, including a plethora of options to pick from, misaligned expectations, unclear navigation, the use of excessive technical jargon, and so on.

In order to reduce bounce rate, it is not possible to apply one solution to all websites because each website serves a distinct purpose and caters to a different audience segment. Running an A/B test, on the other hand, can be advantageous. A/B testing allows you to test numerous variations of an element on your website until you find the one that works best for your audience and business. This not only assists you in identifying friction and visitor pain points, but it also aids in improving the entire experience of your website visitors, encouraging them to spend more time on your site and even converting them into paying customers.

When site visitors find information that they enjoy, they are more likely to spend

more time on your site. User testing to determine the types of content and marketing materials that your users find appealing will assist you in developing a better site — and one that users will want to spend time on.

Improve existing traffic ROI: As the majority of seasoned search engine optimization professionals have discovered, obtaining high-quality traffic to your website comes at a high cost. In order to make the most of your existing traffic and maximize conversions, you should conduct A/B testing. This will save you money because you will not have to spend any more money on getting new traffic. A/B testing can provide a significant return on your investment because, in some cases, even the smallest adjustments to your website can result in a significant improvement in overall business conversions.

Make low-risk modifications: A/B testing allows you to make tiny, incremental changes to your web page rather than having the complete page redone. This can lessen the likelihood of your present conversion rate being jeopardized.

A/B testing allows you to target your resources for maximum output with the least amount of alteration, resulting in a higher return on investment. A good example of this would be a modification in the product description. A/B testing can be used to determine whether or not to remove or update product descriptions on your website. You have no way of knowing how your guests will respond to the change in the environment. By doing an A/B test, you will be able to examine their reaction and determine which side of the weighing scale is more likely to tilt.

Another example of a low-risk adjustment is the addition of a new feature to an existing application. Before adding a new feature, doing an A/B test on it can help you determine whether or not the new modification you're proposing will be well received by your website's target audience.

In the short and long term, implementing a change on your website without first testing it may or may not be beneficial to your business. Testing and then making changes can help to increase the certainty of the outcome.

1.4 TYPES OF A/B TESTS

There are mainly four types of testing methods - A/B testing, Split URL testing, Multivariate testing, and Multipage testing.

Split URL Testing: Split URL testing is an experimentation procedure in which a totally new version of an existing web page URL is examined to determine which one performs better in order to determine which one performs better.

You should use this strategy when you wish to make significant changes to your current page, especially when it comes to the design. You are not willing to make any modifications to the current web page design in order to provide a more accurate comparison.

To establish the winner of a Split URL test, the traffic to your website is divided into two groups: control (the original web page URL) and variations (the new web page URL). The conversion rates of each group are then compared in order to decide which group is the winner.

A/B Testing: You can run A/B Testing on your website when you only want to test changes to the front-end of your website

There are many advantages to doing split URL testing:

• It is strongly advised that you test with non-UI modifications, such as switching to a new database or decreasing the load time of your website.

• Modify the way web page processes are carried out. In business conversions, workflows are very important since they allow for the testing of new pathways before implementing adjustments and the evaluation of whether any sticking spots were overlooked.

• For dynamic materials, there is a better and generally recommended testing procedure that should be used**.**

Multivariate testing (MVT) is a kind of statistical analysis. Multivariate testing (MVT) is a method of experimentation in which changes in a large number of page variables are examined at the same time in order to identify which combination of variables works the best out of all the possible combinations. Unlike a regular A/B test, this one is more extensive, and it is best suited for marketing, product, and development specialists with advanced knowledge and experience. Consider the following scenario: we decide to test three different versions of the pop-up page, each with a different color of the header and picture.

In all, we have nine different versions. The PnC formula is used to determine the number of possible variants., i.e.

(No. of versions of variation A) * (No. of versions of variation B)*…

Advantages of Multivariate testing:

- Reduces the need to conduct many sequential A/B tests with the same goal and saves time because you can track the success of multiple tested page elements at the same time.
- Easily evaluate and determine the contribution of each page element to the measured gains, map all of the interactions between all independent element variations, and calculate the contribution of each page element to the measured gains.

Multipage testing: Multipage testing is a type of experimentation in which you can test modifications to specific items across a number of pages at the same time. A multipage test can be carried out in one of two ways. One option is to take all the pages in your sales funnel and generate fresh versions of each one, resulting in your challenger being the sales funnel, which you can then compare to the control to see which one performs better. This type of testing is referred to as Funnel Multipage testing.

To begin, you can examine how the addition or removal of recurring elements, such as security badges, testimonials, and so on, can affect conversions across the whole funnel. This type of testing is referred to as Classic or Conventional Multipage testing.

Advantages of Multipage testing:

Multipage testing is similar to A/B testing in that it is simple to set up and conduct, and it produces significant and accurate data with ease and in the least amount of time possible.

The following are some of the benefits of multipage testing:

- It gives you the ability to provide consistent experiences for your target audience members.
- Whether they are seeing the control or one of its variations, your target audience will see a consistent set of pages as a result of this strategy.
- You may make the same change on many pages at the same time, which prevents your website, visitors, from being distracted and switching between various versions and designs as they go around your website.

CHAPTER 2

LITERATURE REVIEW

Several studies on A/B testing have been done for different industries that can be used to solve the problems of industries and fulfill their requirements.

"Shafi Kamalbasha and Manuel J. A. Eugster et al." [1] presented "Bayesian A/B Testing for Business Decisions", in which the technique intends to help make data-driven decisions by observing customers' responses and interactions with the interface. They tried to define matters which are easy to understand for managers and marketing managers. Bayesian testing is the easiest format according to them in terms of the A/B test. They demonstrated how they dealt with a few frequent experiment situations that occur in businesses and how they used Bayesian testing to them. They provided the outcome and the ultimate conclusion that was derived from it, all in accordance with the approach. They had the limitation that they observed the multi-option scenario as aggregated scenarios. In the second and third examples, they had ignored many factors considering them as least affecting or not affecting which depends on the company's size, reach, revenue, etc.

"Giordano Tamburrelli and Alessandro Margara et al" [2] created the paper Towards "Automated A/B Testing". This investigation is to automate the A/B testing. By default, A/B testing is a costly, error-prone, and manual activity. They rephrase A/B testing as search-based software's problem and try to automate it using aspect-oriented programming. They contributed significantly to the project in two ways. They ran a series of synthetic trials with their prototype on one hand and real-world experiments with their prototype on the other to demonstrate the usefulness of automated A/B testing. On the other hand, they provided the SBSE community with a one-of-a-kind area in which to use their knowledge and experience. Their future objectives include testing our technique on real users and refining it with specific mutation operators and full support for constraint-based programming.

"Andrea Arcuri and Lionel Briand et al" [3] created the paper "A Hitchhiker's Guide to Statistical Tests for Assessing Randomized Algorithms in Software Engineering".

In this research, the authors performed a systematic study to discover how the results of randomised algorithms in software engineering are investigated and reported their findings. Algorithms of this kind are often used to tackle a broad variety of software engineering problems, including the selection of test cases. They draw the conclusion that the deployment of robust statistical methodologies is insufficient when investigating randomized algorithms in software engineering on the basis of the results, which are consistent with earlier systematic investigations on related subjects. In order to solve this problem, this study explores and defends a set of practical guidelines intended for researchers in the area of software engineering, which are presented in this work. As compared to other scientific fields, the ideas provided in this study are particularly well adapted to the unique characteristics of randomized algorithms when applied to software engineering challenges, as opposed to those found in other scientific domains. Following these guidelines will help to build a credible body of empirical data over time, since they allow for comparisons across studies, which in turn leads to the convergence of generalizable conclusions with practical implications. If the problem of erroneous findings is not addressed promptly, it will have a negative influence on successful technology transfer and will certainly restrict the impact of research on practice in the field of software engineering.

Frank P.A. Coolen et al. [4] published a paper “On some statistical aspects of software testing and reliability”. In this paper, several research challenges are highlighted, and several difficulties are discussed. They conclude that statistical methods cannot provide complete aid for software testing. It is not the goal of this work to develop a new method, but rather to validate existing ones. In many tough software testing scenarios, the assumptions on which most of these approaches are built are shown to be inadequate. Mathematicians and statisticians devised the majority of these methods. In addition to resulting in more efficient software testing and more dependable software, statistical assistance for the software testing process is expected to lead to cost reductions in both the testing process itself and the repercussions of software failures as a result of the statistical support. When working on software system upgrades, it would be especially helpful to have long-term collaboration between the same teams of testers and statisticians. This would apply equally well when working on a range of software systems. Working on a variety of

systems will present significant challenges for testers and statisticians and will likely yield useful additional insights into more generic approaches to statistically supported software testing. Testing upgrades typically benefit greatly from previous experience testing earlier versions, while working on a variety of systems will also provide significant challenges for testers and statisticians.

“Ron Kohavi and Roger Longbotham”, et al., [5] published Online “Controlled Experiments and A/B Testing”. In this paper, they published the structure of the experimentation system and how it can be implemented. Regarding scientific testing methodologies, the Internet and online connection of client software, websites and online services provide a fertile field of possibility for experimentation. It is now widely acknowledged that online experimentation is a critical tool for assessing whether or not a software or design change should be applied in a given situation. It is advantageous to conduct experiments online since it allows you to construct a software platform on which to perform the studies. Consequently, experimentation becomes much more scalable and efficient, allowing for far quicker assessment of concepts.

“Rasmus Ros and Per Runeson, et al”., [6] “Continuous Experimentation and A/B Testing: A Mapping Study”. In this study, a thorough review of the literature on the current status of controlled experiments in continuous experimentation is offered. This review is divided into three parts: On chosen paper sets from two past mapping operations in the field, forward snowballing was applied, and the results were examined. When it came to evaluating the qualitative content of the 128 articles that were finally chosen, thematic analysis was applied. According to the conclusions of the research, a continuous experimentation framework is comprised of two components: an experimental method and a supporting infrastructure that facilitates experimentation. The recommendation to practitioners is based on experience reports that discuss failed experiments in the context of large-scale software development. The recommendation is to implement one of the published processes but to also expand it by putting more emphasis on the ideation phase and by creating prototypes throughout the process. Several research has looked at the infrastructure needs for controlled trials in order to enhance the volume and speed of testing in order to increase the volume and pace of testing. Our proposal for infrastructure will take into

account organizational components to guarantee that, for example, the right channels for communicating results are in place before going on with the project. Data mining, metric specification, variants of controlled experiment design, quasi-experiments, automated controlled experimentation, variability management, continuous monitoring, improved statistical methods, and qualitative feedback were among the ten themes of solutions that were discovered and applied in the various phases of controlled experimentation. During the course of the investigation, six categories of impediments were identified: cultural/organizational issues, commercial challenges, technological challenges, statistical challenges, ethical challenges, and domain-specific challenges.

"Nikhil Bhat and Vivek F. Farias et al". [7] suggested "Near-Optimal A-B Testing", Specifically, they discuss the challenge of A-B testing in situations in which the effect of the therapy is complicated by a high number of confounders. We explore the problem of properly allocating test participants to either treatment with a view to optimizing the precision of our estimate of the treatment impact in such scenarios because randomization can be inefficient in such situations. Their key contribution is a tractable approach for solving this problem in the online setting, where subjects arrive and must be allocated sequentially to covariates selected from an elliptical distribution with the finite second moment, as described in the previous section. They go on to quantify the increase in precision offered by optimized allocations as compared to randomized allocations, and they demonstrate that this increase grows significantly as the number of covariates increases. There are a number of modifications that can be made to our dynamic optimization framework that incorporate significant operational limitations such as the consideration of selection bias, budgets on allocations, and endogenous halting times. After carrying out several numerical tests, they demonstrated that their method provides superior statistical efficiency while also exhibiting less selection bias when compared to competing biased coin designs that are already available.

"Maria Esteller-Cucala, Vicenc Fernandez and Diego Villuendas et al". [8] presented "Evaluating Personalization: The AB Testing Pitfalls" Web Resources That Your Company May Not Be Aware Of—A Look at Websites in the Automotive Industry, This study will discuss several key AB testing dangers that were discovered in the

automobile sector and that may compromise the validity of the studies that have been undertaken in this area. Also included is a study of the occurrence of similar testing concerns in organizations from sectors other than the car industry, which broadens the scope of the analysis. So, we identified five possible pitfall themes and explored their prevalence in the different industries under consideration as a consequence of our findings. Our results reveal that several basic AB testing faults, which are well-known by academics and major digital organisations, are prevalent in the experimental efforts of companies that are relatively inexperienced with AB tests. These flaws include: Four of the research's shortcomings will be discussed in further detail in the next section. Primarily and most importantly, the list of testing risks addressed in this article was first focused on the automobile sector; as a consequence, several crucial dangers for other industries may have gone unnoticed initially. Secondly, this is not an entire list of possible online AB testing issues, but rather a collection of the ones that have been seen and are considered to be the most important, as well as vital, for the overall testing strategy of a corporation. In addition, since the total number of analyzed organizations is insufficient, it is not possible to evaluate the generalizability of each of the testing problems that have been detected statistically. It is possible that further research may widen the scope of the study to cover a greater number of firms. Finally, our investigation is only focused on identifying whether or not certain testing dangers are prevalent across the industry. However, although there is a considerable body of information available on how to detect and avoid these testing issues, the reasons why they may still be identified inside organizations have not been investigated and might be addressed by more study. We think that by publishing this study, we will be able to enhance the degree of awareness among experimenters regarding possible hazards. To further raise the attention of persuasive technology researchers to the discrepancy between academic advancements in the customization area and their application in the marketplace.

Amy Gallo et al. [9] presented A Refresher on A/B Testing. In this article, he explained what is A/B testing. How do we interpret results from these tests, how are companies using this test to grow, and what common mistakes do people make while performing the A/B test.

CHAPTER 3

SYSTEM ANALYSIS

3.1 PROBLEM DEFINITION

To determine which version of a website or app performs better, two versions of the website or app are compared against each other. It is possible to determine which version works better by using a method known as A/B testing (also known as split testing or bucket testing). For website owners, A/B testing, in which several versions of the same live site are shown to users to interact with, is becoming an increasingly popular method of improving their sites. However, although the organization may get valuable feedback and data as a result of this testing, is this the best course of action for the company?

Despite the fact that the findings seem to be black and white, A/B testing still leaves a lot of questions unanswered. While the test designs may be prone to prejudice from those responsible for putting up the test, clients may not get the highest level of service they had hoped for. It also raises the issue of how many testing iterations you will go through and what effect this will have on the goodwill of your consumers' perception of your company.

3.2 ADVANTAGES OF A/B TESTING

Obtain unambiguous proof

It's simple to observe how many more people complete a transaction with site A than with site B when comparing the two sites. In addition, since the evidence is based on real conduct, it offers tangible facts of the kind that money men are interested in learning about (and can be presented in a simple-looking, hard-hitting chart).

Experiment with new ideas and approaches.

An A/B testing procedure will give you actual data as to whether or not your original notion for an existing website will be successful. To put that broad concept through this kind of testing, you will first need to implement it in hard code.

Improve your performance one step at a time by taking little steps

The use of A/B testing is particularly advantageous when dealing with a big site or a large

number of sites since it allows you to "patch" test, starting with a small section of the site and gradually working your way up to its most important parts. Alternatively, can smaller site users with lower traffic bear the risk of giving half of their real visitors with a site experience that may be less than optimal?

Provide explicit design responses to certain design questions

Are green buttons preferable to red buttons when it comes to website design? It is possible to answer these and many other questions using A/B testing since it allows the designer to try multiple colors, button placements, page layouts, and pictures — all of which are ideal places to gradually enhance a website — all at the same time.

3.3 DISADVANTAGES OF A/B TESTING

Takes a significant amount of effort and resources

It is likely that A/B testing will take much longer to set up than other forms of testing. Setting up an A/B system may be a time- and resource-consuming endeavor, but there are third-party services that can help you. The number of meetings held to decide which variables should be included in the tests may be countless, depending on the size of the firm. Design and coding teams must successfully deal with double the quantity of information after a set of variables has been decided upon. Furthermore, for low-traffic sites, studies might take weeks or even months to provide clear findings before they are considered successful.

It only works for specified objectives.

This kind of testing is useful if you want to tackle a specific problem, such as determining which product page produces the best results. However, if your objectives are more difficult to quantify, straight A/B testing will not give you the information you want.

It doesn't make a bad situation any better.

If your site has usability issues, to begin with, and the variants are just iterations of that, it is probable that the variations will still have the core defects that your previous site had as

well as the variations. A/B Testing will not uncover these sorts of flaws or indicate user dissatisfaction, and you will not be able to determine the root cause of the site's issues if you employ this method. Simply because A resulted in more sales does not imply that A is superior than B. It is possible that removing the initial usability problem will be considerably easier to discover and will result in much better outcomes.

It is possible that you may be subjected to frequent testing.

Once the exam is completed, it is the end of the matter. The information is worthless for any other use. Future A/B tests will have to be restarted from the beginning, and other types of testing will most likely only be applied to the more successful site, even if the rejected version had information that was just as relevant.

The most effective use of A/B testing

The use of A/B testing, in conjunction with other testing methodologies, may be a helpful tool in improving a functional design and discovering what attracts your visitors or assists them in completing the tasks on your sites. However, since it is unable to assess ease of use, annoyance, or other factors, it cannot be depended upon as a comprehensive solution. Usability testing should be used to better understand the frustrations and concerns of users, followed by A/B testing to compare, and contrast the many solutions that have been proposed.

3.4 PROPOSED SOLUTION

Because the whole operation is time-consuming, goal-oriented, unrevealing, and continually subjected to human testing, it is very difficult to handle the load on a higher client demand basis than is now required. Because of this, we are focusing on automating the whole procedure. This procedure operates on a response-based basis, responding to the needs of the website/company owner.

Once the process has been automated, the system takes the input and analyses the whole data set, providing each and every exact information that the organization needs to know what adjustments are resulting in more conversions and better results for the company. With automation, there is now the opportunity to modify the length for the firm to watch

the trend, which is especially useful for shorter time periods and fluctuations in the industry. And, since this whole process now just requires that it be launched, it solves our difficulties in that it allows us to save time while also requiring less human work.

3.5 SOFTWARE COMPONENTS

a. Databricks is an industry-leading, cloud-based data engineering platform that is used for processing and manipulating enormous amounts of data, as well as studying the data using machine learning models, to name a few applications. It's the newest big data tool to be brought to the Microsoft cloud, and it was only just launched to Azure.

b. Confluence is a collaborative wiki platform that is designed to assist teams in effectively collaborating and sharing information among themselves. The Team Calendars add-on allows us to manage several calendars at the same time, record project needs, assign tasks to particular users, and capture project requirements in a single place.

c. A member of a family of tools designed to aid teams of all sizes in the management of their work, Jira Software is a good example. When Jira was first developed, it was meant to be used as a bug and issue tracking system. To its credit, Jira has evolved into a powerful task management platform that can be used for a broad variety of applications, from requirements and test case management to agile software development and everything in between.

d. Python is a programming language that is often used in computer programming to develop websites and apps, automate procedures, and do data analysis. Unlike other programming languages, Python is a general-purpose programming language, which means that it may be used to construct a broad variety of diverse applications and is not explicitly developed to tackle any specific problems.

e. When you hear the term "SQL," you are referring to the Structured Query Language, which is a computer language that is used for storing, altering, and retrieving data in relational databases. SQL is sometimes referred to as SQL Server

in certain circles. SQL (Structured Query Language) is the standard programming language for Relational Database Systems (RDBMSs).

f. The Azure cloud computing platform is essentially a vast collection of computers and networking technologies that can be used to run a wide variety of distributed applications. They are in charge of coordinating the configuration and operation of virtualized hardware and software on the servers on which they are placed, as well as the management of the servers. Azure's power comes from the orchestration of these server instances, which is what makes Microsoft Azure so effective.

g. Microsoft Excel is a spreadsheet programme developed by Microsoft that is available for Windows, macOS, Android, and iOS devices. In addition to calculation or computation capabilities, graphing tools, pivot tables, and a calculation or computation capability are included in Visual Basic for Applications (VBFA) (VBA). Excel is a piece of software that is a part of the Microsoft Office suite of programming, which is available for free download.

3.6 DATASET COMPONENTS

Table 3.6 Dataset Components

col_name	data_type	Description
hit_date	date	Recording date
device_part_key	string	Device key
hit_time	timestamp	Recording time
hit_event_id	string	Event Id
device_id	string	Device id
product_analytics_content_version	string	Analysis version
hit_event_type	string	Event type
process_name	string	
__throttle_user_multiplier	int	
hit_type	string	
hit_category_0	string	Mesg id
hit_category_1	string	Misc info
hit_category_2	string	
hit_screen	string	
hit_action	string	Event action
hit_label_0	string	
hit_label_1	string	

hit_label_2	string	
hit_label_3	string	
hit_label_4	string	
hit_feature	string	
hit_trigger	string	
hit_source	string	Software name
hit_severity	string	
hit_guid	string	Msg_id
hit_engagement_desired	string	
hit_session_id	string	
hit_engagement_interactive	string	
hit_engagement_userinitiated	string	
hit_result	string	Action of event
hit_reserved_1	string	
hit_reserved_2	string	Misc info
hit_reserved_3	string	
hit_reserved_4	string	
hit_reserved_5	string	
hit_metric_0	int	
hit_metric_1	int	
hit_metric_2	int	
hit_metric_3	int	Id for infos
hit_metric_4	int	
hit_metric_5	int	
hit_metric_6	int	
hit_duration_seconds	int	Event type
hit_size_inbytes	int	
device_cpu_model_description	string	Device_details
device_disk_count	int	
device_disk_primary_encrypted	string	
device_disk_primary_size_available_gb	int	
device_disk_primary_size_total_gb	int	
device_disk_primary_type	string	
device_manufacturer_id	string	
device_manufacturer_model	string	Device model
device_manufacturer_name	string	
device_os_install_date	date	Os install date
device_os_machine_guid	string	
device_os_timezone_bias_hours	int	
device_os_timezone_name	string	
device_os_type	string	Windows/ios
device_os_version	string	Os version
device_os_wsc_antimalware_enabled	string	
device_os_wsc_firewall_enabled	string	
device_os_wsc_protection_status	string	

device_ram_size_mb	int	
device_screen_resolution	string	
product_alertmanager_version	string	
product_analytics_bootstrapfile_version	string	
product_analytics_version	string	
product_csp_client_id	string	Client_id
product_csp_policy_ngm_message_preuninstall_enabled	string	
product_csp_policy_ngm_message_registration_enabled	string	
product_csp_policy_ngm_message_subscriptionmilestone_enabled	string	
product_csp_policy_ngm_scheduledtask_enabled	string	
product_csp_version	string	
product_modcore_version	string	
product_msad_uuid	string	
product_msad_version	string	
product_mss_version	string	
product_ngm_cohort_id	string	Cohort_id
product_ngm_enabled	string	
product_oem_entitlement_complete	string	
product_oem_ipr_done	string	
product_oem_jeula_status	string	
product_oem_oc_sku	string	
product_sustainability_version	string	
product_wss_affiliate_id	string	Aff_id
product_wss_amcore_version	string	
product_wss_cohort_legacy_id	string	
product_wss_coreui_activationscreen_version	string	
product_wss_coreui_enabled	string	
product_wss_coreui_module_version	string	
product_wss_coreui_version	string	
product_wss_culture_id	string	
product_wss_eula_date	string	
product_wss_eula_state	string	
product_wss_hardware_id	string	
product_wss_install_date	string	Software_install date
product_wss_install_factory	string	
product_wss_install_type	string	
product_wss_mpf_version	string	
product_wss_mps_version	string	
product_wss_mqs_version	string	
product_wss_msc_version	string	
product_wss_msk_content_version	string	
product_wss_msk_version	string	
product_wss_ngm_version	string	
product_wss_reboot_pending	string	

product_wss_release_internal_name	string	
product_wss_release_market_name	string	
product_wss_reseller_id	string	
product_wss_software_id	string	
product_wss_subscription_account_id	string	Subscription details
product_wss_subscription_affiliate_id	string	“
product_wss_subscription_applist	string	“
product_wss_subscription_eula_set	string	“
product_wss_subscription_expiry_date	date	“
product_wss_subscription_expiry_remaining_days	int	“
product_wss_subscription_extendedexpiry_date	date	“
product_wss_subscription_extendedexpiry_remaining_days	int	“
product_wss_subscription_flex_package_id	string	“
product_wss_subscription_key	string	imp
product_wss_subscription_package_id	string	“
product_wss_subscription_registered	string	“
product_wss_subscription_segment_id	string	“
product_wss_subscription_segment_type_config_id	string	“
product_wss_subscription_segment_type_id	string	“
product_wss_subscription_trial	string	“
product_wss_systray_icon	string	
product_wss_vscore_version	string	
product_wss_vso_canrun	string	
product_wss_vso_content_version	string	
product_wss_vso_version	string	
product_wss_vul_version	string	
product_wss_vpn_integration	string	Vpn detail
product_wss_vpn_deep_integration	string	
product_wss_vpn_version	string	
product_wss_coreui_visible_apps	string	
product_wss_coreui_entitled_apps	string	
product_wss_coreui_installed_apps	string	
product_wss_vso_lam_fileless_enabled	string	
device_geo_id	string	
device_env_qa	boolean	
product_wss_oem_oobe_complete	boolean	
product_wss_coreui_ux_id	string	
product_vantage_version	string	
device_os_oobe_complete_time	timestamp	
hit_cumulative_device_uptime_seconds	int	
hit_cumulative_internet_uptime_seconds	int	
product_analytics_datasets_catalog_version	string	
product_analytics_da_definitions_version	string	
product_analytics_dictionary_version	string	Analytics data

product_analytics_events_version	string	
product_analytics_profile_version	string	
product_analytics_data_items_version	string	
device_vpn_available_networks	string	Vpn networks
device_vpn_active_network	string	
context_reserved_0	string	
context_reserved_1	string	
context_reserved_2	string	
context_reserved_3	string	
context_reserved_4	string	
context_reserved_5	string	
context_reserved_6	string	
context_reserved_7	string	
context_reserved_8	string	
context_reserved_9	string	
stream_date	string	Streaming date
stream_hour	int	And time
product_wss_coreui_enabled_apps	string	
product_wss_coreui_disabled_apps	string	
product_wss_coreui_setup_apps	string	

3.7 USE CASES

A/B tests can be used in different types of industries under different circumstances and for different outputs. It can be also used for different durations of tests and in some cases test duration can be modified during the test, although it effects the accuracy of the test also increase the risk of false results

A/B testing on Mobile elements

When it comes to marketing efforts, A/B testing, also known as split testing, is well-known among professionals as a vital tool for making better, more data-driven choices. Among other things, marketing professionals use A/B testing to test ad content and creativity, email subject lines, Web sites, and mobile apps the most regularly.

A/B testing Facebook adverts and email subject lines is something that practically every marketer is doing these days, and it is becoming more popular.

For mobile marketers, A/B testing refers to the process of refining conversion funnels inside your mobile app by experimenting with alternative calls to action and new user flows, as well as different graphics and layouts, to see which ones perform better. Advertising that increases

mobile acquisition is often subjected to A/B testing.

While A/B testing solutions may be used for basic optimization, they can also be used for a variety of additional purposes.
Here's an example of A/B testing that is commonly overlooked yet is really beneficial, especially for mobile marketers: split testing.

Target messaging

In today's world, one of the most difficult challenges for business intelligence experts is the sheer volume of data available, making it more difficult to act on the data than it is to obtain or, in some situations, even comprehend it.
As clients become more flooded with marketing messages from all angles, creating highly tailored experiences is critical to distinguishing oneself from the competition. This is another issue that A/B testing may aid in tackling with the use of split testing. Most A/B testing systems have other capabilities in addition to the ability to send version A to a random 50 percent of users and version B to another random 50 percent of users, which is standard practice.
With the use of a mature A/B testing solution on the Web or on mobile, a marketer may define targeting criteria based on a range of unique elements, such as device, operating system (including mobile), region, language, new vs returning customers, and any number of other bespoke characteristics.
Marketers can now reach a broader audience more successfully than ever before because of the ability to tailor campaigns to individual users based on demographic, regional, and behavioral information collected from their websites.
Since 2012, advertisers have been able to target individuals based on their demographic information on Facebook's platform. Our mobile application or Web site needs the same degree of complexity that we have applied to our desktop applications and websites.
The ability to compare data from your focused message to data from your untargeted message is an extra benefit of A/B testing, which helps you to figure out how much money you are really making on your marketing expenditure.

A/B Tests for Different Business- Validating what customers want

When clients arrive on a web page or any other piece of marketing material from your firm, they should get the impression that the organization understands their demands. Jerry Han, chief marketing officer at Prize Rebel, concurs with this assessment. The ability to illustrate the value of your product and communicate precisely why a user should take a particular action is what effective messaging is all about, according to him. It is based on well-researched ideas about what would convert lookers into purchasers that effective copywriting is built upon. Moreover, the purpose of testing is to determine whether or not these hypotheses are right.

If you see that your conversion rates or impressions are lower than you would expect, Han suggests that your website copy could be the cause of the problem. A/B testing your website copy can assist you in determining whether or not you can improve the messaging aspects on the page. When it comes to copy experimentation, it is not enough to change a few words in order to try to influence a user subconsciously. Experimentation and copywriting are effective when used together since brands have no way of knowing the true benefits of a product or service unless they test it out for themselves. In order to increase clickthrough rates, it is necessary to test alternative storylines that highlight different product benefits. It is feasible to double or even triple clickthrough rates by experimenting with different narratives.

Right Keywords For Pay-Per-Click(PPC) Campaigns

Because the majority of PPC advertisements are powered by your keyword selection, it makes sense to optimize for the most relevant keywords for your target audience, even if you are running multiple campaigns at the same time.

Home Buyer Louisiana's CEO and creator, Stephen Keighery, explains that because their campaigns run for long periods of time, it is critical that they get the most out of their budget. In order to conduct A/B testing, they create many campaigns, each with a different set of keywords, and then run them all at the same time over the duration of six months. The outcomes of which are used to shape their future campaign as well as other marketing efforts.

Absolute Reg LTD is one of the thousands of businesses that utilize sponsored Facebook advertisements to promote their products and services, as well as A/B testing to determine which campaigns are more effective in terms of return on investment.

"After the [test period] is through and the results come in, we examine the return on investment (ROI) of each campaign in order to identify which is the most effective for our target audience," says Jake Smith, managing director. "After that, it's just a matter of copying the most successful

campaigns and updating them as needed."

So, how do you go about choosing the most appropriate keywords for your digital marketing campaigns? When it comes to selecting the most appropriate keywords for PPC advertisements, Google explains it best: "The most appropriate keywords can put your ad in front of the most appropriate clients." Their user-friendly Google Ads Keyword Planner not only assists you in selecting the most appropriate keywords to target for your campaign but also gives you bid estimations for those particular keywords. Do you prefer to spend X dollars on a high-volume search phrase or Y dollars on a lower-volume but more specialized keyword?

This gets us to the question of how you might incorporate A/B testing into your keyword selection process. In order to determine which keywords generate the most qualified leads, the first step is to test the keywords themselves. The second strategy is to test different versions of the same advertising (as previously discussed) but for different keywords in order to find the most effective combination.

A/B Tests on HealthCare

A banner ad is a digital advertisement that appears on third-party websites when you think about banner adverts. However, this is not always the case. You can develop banner advertisements for your own website in order to promote your own campaigns.

In the example of Humara, a healthcare insurance provider based in the United States, they were experiencing poor participation in one of their initiatives that provided social media insights to those working in the healthcare profession.

Following extensive testing, they discovered that the landing page without a banner received the best response from their target population. Developing a hypothesis and testing it against the control, as they did in the previous case, was necessary to determine this (the original).

This is consistent with the findings of the Forrester Opportunity Snapshot, which concluded that healthcare organizations (HCOs) who adopt a digital-first marketing strategy outperform their competitors in terms of customer experience (CX).

Leaders in healthcare organizations who are responsible for marketing and customer experience (CX) understand that data, analytics, and the insights that flow from them are critical in ensuring that healthcare customers have great experiences.

In a recent survey, 72 percent of healthcare firms reported that they employ A/B testing as a personalization approach to constantly improve and optimize experiences for their clients and patients.

According to the Forrester report, "Managers should rely on profiles and recommendations to make the most of each client visit, as well as testing to determine why a specific encounter is useful."

Personalization can be driven by HCOs through the use of segmentation and data, and sophisticated testing capabilities can be achieved through the use of personalization software. It is possible to run A/B testing and other types of personalisation while still complying with the Health Insurance Portability and Accountability Act (HIPAA) if you are worried about remaining in compliance with the law.

CHAPTER 4

SYSTEM DESIGN

4.1 SYSTEM ARCHITECTURE USED

The following are the components of an experimental system: The fundamental arrangement is to analyze an asset on two different levels: control and test (or test and control). A control version is the default case, whereas a test version is one in which the variation has been tested and evaluated. In the business, this test is referred to as A/B/n split testing since it is more comprehensive. In scientific literature, a multifactor experiment is referred to as a multivariable or multivariate experiment, depending on the context.

The high-level framework of an A/B experiment may be understood in the manner shown in the figure. The test and control may be given any proportion in a practical setting; however, 50 percent offers the most statistical power for the test, and it is recommended to power the trials at lower percent ages after the ramp-up phase to check for major mistakes.

In general, the study looks at whether the statical distribution differs between the test and control conditions. In practice, we compare the two means to see whether they are equivalent. For the same reason, the test's impact is defined as follows:

E(B)= X’B - X'A

The mean of the test is symbolized by the symbol X'B, where X is a measure that is considered to be significant. The percentage change is shown in combination with an appropriate time period to illustrate the relationship. In every control system, the two most crucial components to consider are extraneous influences and unpredictability, respectively. It is possible to categorize everything that has an effect on the test as either an experiment factor or a non-experiment component, depending on how it impacts the examination process. Depending on the scenario, the non-experiment component may be kept constant, blocked, or allocated at random to a different condition. Keep a constant for a factor since doing so may have an effect on external validity and is thus not advised.

If you know that weekend days are different from weekdays, you may conduct the

experiment exclusively on weekdays; nevertheless, it is preferable to conduct the experiment over a period of many weeks to guarantee stronger external validity. Using blocks rather than randomization may assist to limit the amount of volatility in a dataset. In particular, when the experimental units inside each block are more homogeneous than the experimental units between blocks, blocking is beneficial.

When the randomization unit is a user page visit, it is feasible that blocking by weekend/weekday will minimise variation in the effect size, resulting in improved sensitivity. It is critical to randomize throughout the experiment since many external variables change over time. To achieve this, run the control and test(s) simultaneously with a set proportion to each throughout the experiment with a set %age to each. Containing a non-treatment component assures that it has an equal impact on both the control and the test, and as a result, has no effect on the estimation of the test effect.

The hypothesis is shown in Fig. 4.1.

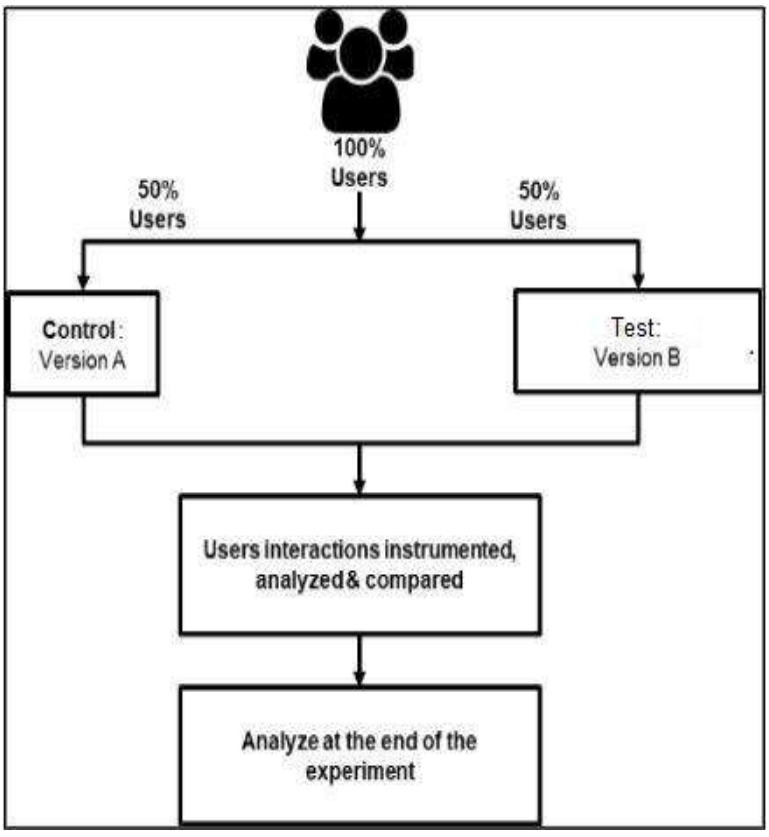


Fig 4.1 System Architecture

4.2 ARCHITECTURE DIAGRAM WORKING

Figure 4.2 shows an architectural diagram.

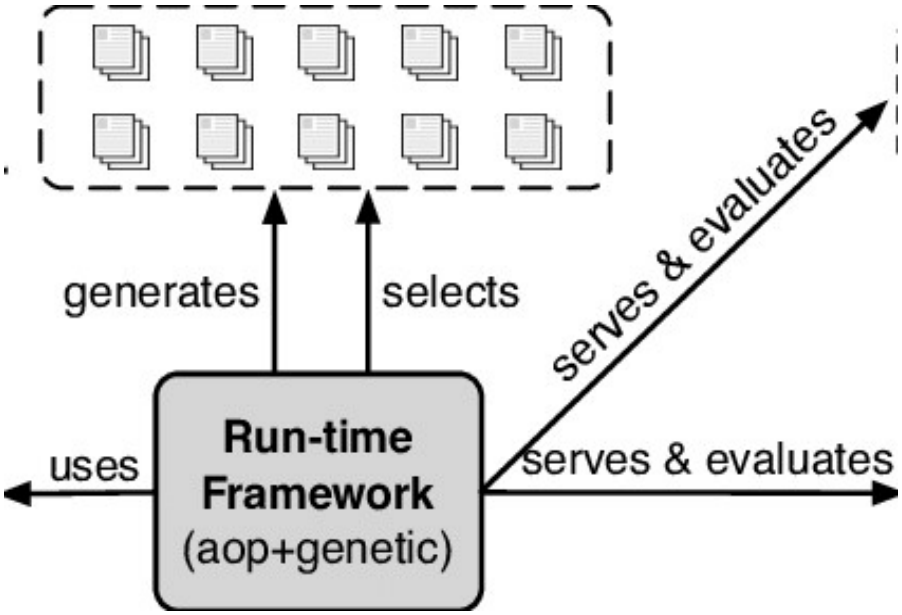


Fig 4.2 Working Architecture

4.3 CALCULATE SAMPLE SIZE USING MDE

When the experiment is completed, you should expect to observe the lowest feasible improvement in conversion rate over the conversion rate of the current asset (the baseline conversion rate) as a consequence of the experiment.

By customizing MDE, you may determine the amount of conversion rate increase that will be necessary for the algorithm to declare the new asset as the winner of the competition. As you decrease the MDE number, the more subtle the conversion modifications that will be observed by the system will be, and vice versa. An experiment's sensitivity may be evaluated using the MDE approach, which is in essence a sensitivity evaluation technique. Having a large sample size is related to having high sensitivity characteristics, such as a low MDE. In general, the lower the MDE, the larger the volume of traffic necessary to detect even minor changes, as well as the amount of money required to generate that traffic.

You will have more freedom in matching the experiment design with the expenditures that you are prepared to endure as a result of developing MDE.

4.3.1 THE EFFECT OF MDE ON THE SAMPLE SIZE

The use of MDE has a considerable influence on the amount of traffic required to reach statistical significance in research when it is used. When doing sequential A/B sampling, you may use the Evan Miller calculator to determine the maximum sample size that can be used. Make sure that the MDE value is entered as a relative value rather than an absolute number when entering the MDE value. By choosing a lower MDE, you teach the system to detect slighter conversion rate changes, which necessitates more traffic and, maybe, more time on the system. Alternatively, the bigger the MDE you specify, the less traffic (and, as a result, the shorter the time required to complete the test) will be generated. Please bear in mind, however, that the sample size you see is only the upper limit of what is required to get a statistically significant result in this particular instance. Because of the nature of sequential A/B testing, the system will continuously investigate the difference between the conversion rates of the variations under consideration. Identifying and scoring the difference is sufficient after the test is completed; it is not essential to score the whole sample set.

4.3.2 FUNCTIONAL BLOCK DIAGRAM

Figure 4.3 shows the functional block diagram.

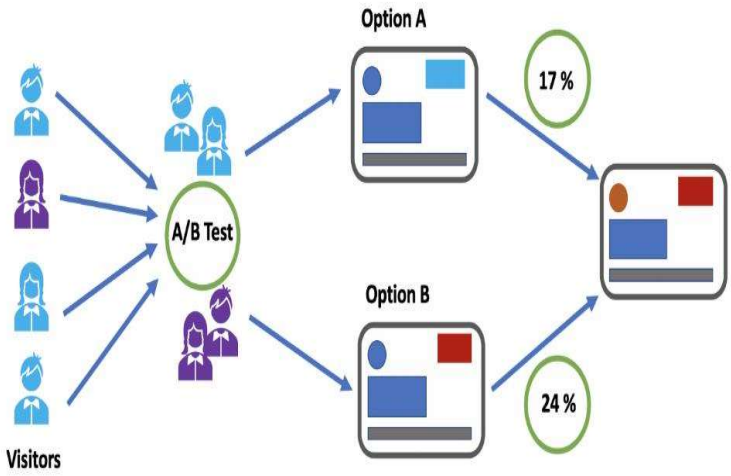


Fig 4.3 Block diagram

4.3.3 HOW TO CALCULATE MDE?

Despite the fact that our system can count MDE for you, we highly encourage that you do it manually. This parameter is dependent on your own risks – the amount of money you are willing to spend on traffic acquisition and the amount of time you are willing to wait for the experiment to run. You must comprehend the following concepts in order to get MDE that is effective for you:

- Traffic acquisition expenses, or the money you spend on generating the appropriate sample size;
- potential revenue produced from ASO-acquired users, or the money you stand to gain by employing the new asset at a better conversion rate;
- and potential revenue generated from ASO-acquired users.

The highest feasible MDE indicates that the prospective income surpasses or completely compensates for the expenses of traffic collection and distribution.

4.3.4 WORKFLOW FOR MDE COMPUTATION

Step 1: Calculate the needed increase in conversion rate.

If your product page's conversion rate with the present icon is 20%, you should consider changing it (baseline conversion rate). You make the assumption that the new icon should have a conversion rate of at least 22 % in order for you to choose it over the present icon. As a result, you must design an experiment in such a manner that it announces the winner when the difference in the conversion rates is at least 22 % – 20 % = 2 % (or more). In order to do so, you must first calculate your expected MDE.

MDE is expressed as a percentage of the baseline conversion rate, which is as follows: MDE is equal to the intended conversion rate lift divided by the baseline conversion rate multiplied by 100 %. In this example, 2 % of the baseline conversion rate of 20 % is 10 % - this is your anticipated mean difference between the two groups for the trial.

Step 2: Calculate the number of participants in your study.

The next step is to calculate the sample size for sequential A/B testing, which may be done using Evan Miller's calculator. Calculation of the sample size for studies with two modifications (A and B).

Listed below is the information you should provide into the calculator:

- Your anticipated Minimum Detectable Effect is 10 %, according to your calculations (in this example). Important! Check to verify that you're using relative MDE.
- Fill in the "Baseline conversion rate" area with whatever number you like. Because we are using relative MDE, the baseline conversion rate is not taken into consideration when calculating the sample size. In Split Metrics, statistical power is set to 80 % by default; the significance threshold is set at 5 % (default in Split Metrics).

The following will be displayed:

- If there are 2,922 total conversions – this is the maximum sample size per two variants (A+B) required to complete the experiment – control wins the trial.
- This indicates that the system will successively check the difference in conversions between variation A (control) and variation B, and may complete the experiment if the difference of 106 conversions is detected, even if the maximum sample size has not yet been reached.

For studies with numerous variants (A+B+C+...), the sample size calculation is performed. Listed below is the information you should provide to the calculator:

- MDE is an estimate. It is OK to use any amount as the baseline conversion rate.
- 80 % of the population has statistical power.
- Statistical significance level: When more than two variants are examined, the Sidak correction must be applied to the statistical significance level.
- Each pair of variants has a unique amount of importance associated with it. In proportion to the number of variables under investigation, the overall significance level increases as a result of the accumulation of individual values.
- The Sidak adjustment equalizes individual significance levels such that the total significance level is equal to 5 % after the correction is applied.

Step 3: Calculate the total number of conversions.

After you have entered all of the information above into the calculator, the total conversions will show. For example, you could wish to conduct an experiment with three different versions - A, B, and C. The following are the items you'll need to enter into the calculator:

- 10 percent (it could be a different percentage depending on the conversion rate lift you want the system to detect); estimated MDE: 10 percent (it could be a different percentage depending on the conversion rate lift you want the system to detect); estimated MDE: 10 percent (it could be a different percentage depending on the conversion rate lift you want the system to detect);
- 80 percent of the population has statistical power.
- The significance threshold is three percent.
- The total number of conversions necessary is 3,472.
- Multiply the total number of conversions by two.

You'll receive the following results if you use the previous example with three variations:

- 3,472 divided by two equals 1,736 divided by two equals 3,472 divided by two equals 1,536
- Increase the value acquired in step 3 by the number of versions, which includes the control, to reach the following result:
- The formula for A+B+C is as follows: total conversions / 2 *3
- It is necessary to divide the total number of conversions by two and four for A+B+C+D. Recall that if you are doing an A+B+C experiment, the multiplier will be three: A, B, and C
- For A+B+C+D+E, the following formula is used: total conversions / 2 * 5 = total conversions for A+B+C+D+E.

Recall that if you are doing an A+B+C experiment, the multiplier will be three: A, B, and C

- The sum of all conversions multiplied by three times two equals 5,208.
- The estimated maximum sample size for an experiment with three variations (A+B+C) is 5,208, which is a cautious estimate based on previous experience.
- After determining the maximum sample size necessary, you may assess the potential expenses associated with traffic acquisition. Make use of the following formula:

Traffic acquisition expenses are calculated as follows:

“total conversions/baseline conversion rate * cost per click.”

You may calculate the number of visits in your sample by dividing the total number of conversions by your baseline conversion rate (those who click on your ad banner). Consider the following scenario: Your cost per click is $0.5, and your conversion rate is 20 percent at the start of the campaign. (Convert it to decimal form so that it may be used in the formula). The following will be your traffic acquisition costs:

$2,922 /0.2 * $0.5 = $7,305.

It is possible to specify "Complete Registration" as a conversion event when using Split Metrics in conjunction with the Facebook Pixel integration. Instead of counting visitors who click on an ad banner, the traffic acquisition expenses will be determined based on users who click on the "Get" button in this scenario. In such circumstances, the following will be included in the cost calculation formula: “Cost per Install (not Cost per Click)”

Traffic acquisition expenses are calculated as follows:

“total conversions * CPI”

In the event that additional visits are made, there is no need to change the sample size. Simply multiply the Consumer Price Index (CPI) by the total number of conversions made possible by the Evan Miller calculator and you have the answer. In the example above, if your cost per install is $2.5 and your maximum sample size is 2,922 total conversions, your cost per install would be $2.5 per install. The following will be your traffic acquisition costs:

$2,922 * $2.5 = $7,305.

Examine whether or not these expenditures are in agreement with the budget that has been set aside for traffic acquisition at this point by looking for the following signs of compliance:

Consider increasing the MDE whenever possible since a bigger sample size will result in cheaper acquisition expenses. If your variants are highly similar to one another, a high MDE score will not be able to deliver a relevant conclusion. You may continue to make educated guesses about the prospective earnings of ASO-acquired users.

Step 4: Make a list of all of the things you want to do. Make an educated estimate about the amount of potential revenue.

You may calculate the potential revenue from the conversion rate lift using a number of approaches, including the lifetime value (LTV) of app subscribers obtained with ASO, among others. When dealing with two versions (A and B), it is required to determine the amount of money that will be created by a 2 percent increase in the conversion rate in the example above.

The following table shows how you might compare your prospective revenue ($Y) to your traffic acquisition expenditures ($X) after you've established your potential income ($Y).

The estimated MDE in your experiment may be used if the prospective revenue surpasses the traffic acquisition costs ($Y > $X); but, if the prospective revenue exceeds the traffic acquisition costs ($Y $X), you must estimate a bigger MDE and rerun all of the steps outlined above.

CHAPTER 5

SYSTEM IMPLEMENTATION

5.1 ALGORITHM USED

In terms of particular information, such as formulas and notions, we refer to Gelman's "Bayesian Data Analysis." It is founded on the notion of probability assertions that Bayesian statistics operates on. It also explains how to change the probability if new information has been obtained. Confinacy is a term used in Bayesian statistics to indicate that the posterior probability distribution P(A|X) belongs to the same probability distribution family as the prior probability distribution P. Confinacy is defined as the probability that the posterior probability distribution P(A|X) belongs to the same probability distribution family as the prior probability distribution P. In other words, the posterior distribution P(A|X) belongs to the same probability distribution family as the prior distribution P(A|X) (A). This kind of prior is referred to as a conjugate prior in this instance.

$$P(A|X) = \frac{P(A)P(X|A)}{P(X).}$$

The link between the posterior distribution P(A|X) of parameter A after viewing data X and the previous distribution P(A) of parameter A is described by the posterior distribution function. This model necessitates the use of the following six distributions: the binomial distribution, the Beta distribution, the multinomial distribution, the Dirichlet distribution, the exponential distribution, and the Gamma distribution, among others. The binomial distribution is the most often used distribution in finance, accounting, and economics.

We have several different experiment scenarios, such as comparing variations where the end-user has only one option to choose from, comparing variations where the end-user has many options to choose from, and comparing variations with many options but only observing the final lift as a result of the final lift. Our team will be working on the third model, which is then aggregated model, in this instance.

5.1.1 AGGREGATE MODEL

The aggregated conversion rate as well as the sum of income creation per option are observed during this procedure. This occurs as a result of an excessive number of alternatives inside a variant or as a result of the data creation procedure. Some of our tests have more than 20 different possibilities to pick from, and we formalise the scenario as follows to ensure that all essential requirements are met.

For a given variant a, there are K_a unknown options. N_a is total visitors and C_a is total conversions. The individual revenue per instance V_a is unknown. Executing the experiment results in collected data C_a, and the aggregated revenue S_a overall C_a successes and implicitly over all unknown K_a options, i.e., $S_a = \sum C^i_{j=1} S_{ij}$ with S_{ij} the revenue of each success j. To get an estimate of the revenue per visitor V_i we model the average revenue per visitor V_i' given the observed aggregated revenue S_i. For that, we assume that S_{ij} follows an exponential distribution

$$P(S_{ij}|V_{i'}) = \text{Exp}(V_{i'})$$

where V_i' is the scale parameter of the distribution. The exponential distribution means that lower revenue has more probability of occurrence than higher revenue. This assumption fits our observation of the visitors' money spent curve on our website. The prior distribution of V_i' is modeled as

$$P(V_i) = \text{Gamma}\ (\alpha_i\ ,\ \beta_i)$$

Taking advantage of the conjugacy relationship between exponential and gamma distributions, we compute the posterior distribution as

$$P(V_i|S_i) = \text{Gamma}\ (\alpha + C_i\ ,\ \beta_i + S_i)$$

The expected value per client per variation is $A_i * V_i$. Since we have the posterior distributions for both, A_i and V_i, we take n random samples from $P(A_i|X_i)$ for the conversion rate Y_i, n random samples from $P(V_i|S_i)$ for the revenue.

5.1.2 MAKING DECISIONS FOR THE TEST

For each of the tests, a feature has been defined, which determines whether the test was successful or unsuccessful, or which variation will produce a better result in terms of conversion rates, uninstallation rates, click-through rates, and other metrics.

5.1.3 PROBABILITY OF GETTING THE BEST RESULT:

After running a test with many variations, we want to derive the calculation for the best variation and which one to be implemented. Given a set of posterior samples Y_i of the measure of interest from *i* different variants, the probability for the best is defined as the probability that a variation has a higher measure in comparison with all other variations. The probability that Y_1 is better than Y_2 is the mean of:

$$P(Y_1 > Y_2) = [y_{1j} > y_{2j} \mid i \text{ and } j \in \{1, \ldots, n\}]$$

To find the best variation, we analyze all combinations and select the one with the highest probability.

5.1.4 EXPECTED LIFT

After getting the best variant, we also need to check the increase in the measure of the matrix we expect. Given a set of posterior samples Y_i, the expected uplift of choosing Y_1 over Y_2 is defined as the mean of the percentage increase:

$$U(Y_1, Y_2) = [(y_{1i} - y_{2i})/ y_{2i} \mid i \text{ and } j \in \{1, \ldots, n\}]$$

We want uplift compared to control variation only, we calculate it for all test variations against each other.

5.1.5 LOSS THAT IS EXPECTED

Consider the following scenario: Variation 1 has the greatest likelihood of being the best, yet it is less than 1. There is still a potential that the other variety will turn out to be the genuine top performer in the end. In such a circumstance, we would want to know the risks associated with selecting Variation 1. Given a set of posterior samples Y_i, the expected loss when choosing Y_1 over Y_2 is the mean of:

$$L(Y_1, Y_2) = [\max((y_{2j} - y_{1j})/y_{1j}, 0) \mid i \text{ and } j \in \{1, .., n\}]$$

Like the expected uplift, the expected loss is calculated between the test variant and the control.

5.2 MATHEMATICAL IMPLEMENTATION

Click-Through Rate (CTR):

It is for usage purposes that the number of total views or sessions is taken into consideration; This percentage represents the number of individuals who view the page (make an impression) before actually clicking on it (clicks).

$$CTR = \frac{\#\ total\ clicks * 100\%}{\#\ total\ clicks + \#\ total\ views}$$

Click-Through Probability (CTP) :

Unlike the click-through rate (CTR), the CTP takes into account duplicate clicks. This means that if a user for some reason performs multiple clicks on the same item in a single session for some reason (for example, because of impatience), these multiple clicks are counted as a single click in the CTP.

$$CTP = \frac{\#\ people\ with\ at\ least\ 1\ click * 100\%}{\#\ number\ of\ unique\ visitors\ per\ page}$$

5.3 EXPERIMENTAL SETUP

Formulate a hypothesis

Before you begin testing the other assumptions, decide which assumption you want to test first, and then test that assumption. In order to do this, you must express your theory as accurately as possible. They should not be contrived, and it is critical that they have a good logic to support them in their use. Using web analytics and an assessment of visitors' behaviour on your website or other online resources, develop hypotheses about your target audience.

When everything goes according to plan, the following hypothesis develops as a consequence of the investigation: Customers should be asked why they like your product and what issue it addresses for them, according to the company. Analyze support requests to discover where the bottlenecks are located. Take a look at the figures: Search your site for pages with a high proportion of outputs or pages with an unusually low conversion rate, and use the analytics to keep track of what people are doing on your site on a regular basis and how they are behaving.

For example, online analytics show that customers seldom click on the "Buy" button, which is shown on the website. Possibly as a result of her self-centered nature? Is it because it blends in with the general backdrop why it is so popular? While it is possible to develop numerous hypotheses, you can only test one of them at a time in a single experiment. That is, two distinct tests are necessary when evaluating a button: one to determine the button's size and another to determine its colour. If you test a large number of hypotheses at the same time, you will not be able to tell which ones were successful in your experiment.

One important criteria to remember is that only one modification may be examined in a single experiment, and this is a critical consideration.

Recognize and describe variances

As soon as you've decided which hypothesis you'd want to test, the following step is to determine the criteria that will be used for that specific test.

In addition to the bounce rate (the number of visitors who visited a page and then immediately departed), time spent on service resources, the number of

applications or registrations, the number of transactional transactions, and the average check amount may all be considered as criteria.

When we go back to the beginning, we need to figure out what the issue is that needs to be resolved first. You might, for example, offer an online course on self-pleasure to get money. You are attempting to increase sales for one of the three programmes, and you want to bring attention to it by highlighting it in a different colour.

Select one variation

This is what you'll use to put your theory to the test. In the case of a red button, for example, if you believe that it will be more visible on the page and will boost the number of registrations, you must test the colour. You can test any element on your page, including the following:

- Content and location of headings and subheadings.
- Calls to action (length, content, location).
- Buttons (colour, size, location, text).
- Images (size, content, location).
- Text on the page (length and content).
- forms (location, size, and number of fields).
- Prices of goods and services

After you've decided which element you'll be testing, you'll need to make a copy of the page and make the necessary adjustments. Furthermore, as previously stated, you may only test one hypothesis at a time and make one change during a single test.

Calculate the sample size for the test

In order to accomplish this, you must first determine how many people must visit the website in order for the test findings to be statistically significant (that is, they could not have occurred by chance) and reliable.

The sample size is determined by how much of a difference you expect to see. The greater the impact of these changes, the fewer people will be required to

participate in the sampling. As an example, you can compute it with the assistance of internet calculators such as Optiizely, A/B Tasty, Unbounce, A/B Test Guide, and others. We can also calculate the sample size using MDE.

Consider if you want to show the pilot page to all visitors or only to a subset of them before proceeding. If you show everyone, the required number of people will be reached much more quickly. Alternatively, if the theory is proven incorrect and the new page has a lower conversion rate, the company will suffer a financial loss.

Determine the duration of the experiment

Even if you had the correct amount of visits in two days, the test should last at least a week. This is required due to the fact that users react in various ways on different days of the week. If the target signal is a purchase, and you are aware that most customers do not make a buy right away, but rather wait 10 days, then this should be taken into consideration. The usual amount of time recommended for testing is 10–14 days. Once a time has been set, never interrupt the test before it has finished, even if one choice appears to be clearly in the lead during the first few days. It takes time to achieve solid outcomes.

Do A/A testing

This is the process by which you offer all users two pages that are completely identical and then track the outcomes. The findings of A/B testing will be reliable if they are the same for both. If they are different, then the traffic is not uniform. If the results of one page in an A/B test are significantly different from the results of another page, then it is not necessary to run an A/B experiment because the results will be automatically wrong.

A/A testing can be done out within the primary test to save time, resulting in A/B/A testing being performed as a result. In order to accomplish this, you must divide the traffic into three similar streams and present option B to one of them while presenting option A to the other two. It is reliable if the results of pages A and A coincide, which indicates that the exam was fair.

Because of this, you would need double the amount of traffic compared to a

typical technique in order to produce statistically significant results with this strategy.

Do A/B Testing

The primary test may be conducted when you have formulated a hypothesis, defined objectives, determined the length and sample size, and confirmed that the traffic is uniform. Specialized service, such as Google Experiments, Change Again, VWO, Optimizely, or AB Tasty will be required in order to carry out the experiment successfully. All of these services are diverse from one another in terms of pricing as well as function. Start by familiarizing yourself with Google Trials and doing your experiments there, since this will save you time later on. It is completely free and has a direct relationship with Google Analytics. Steps to perform the test have been described in Fig 5.3.

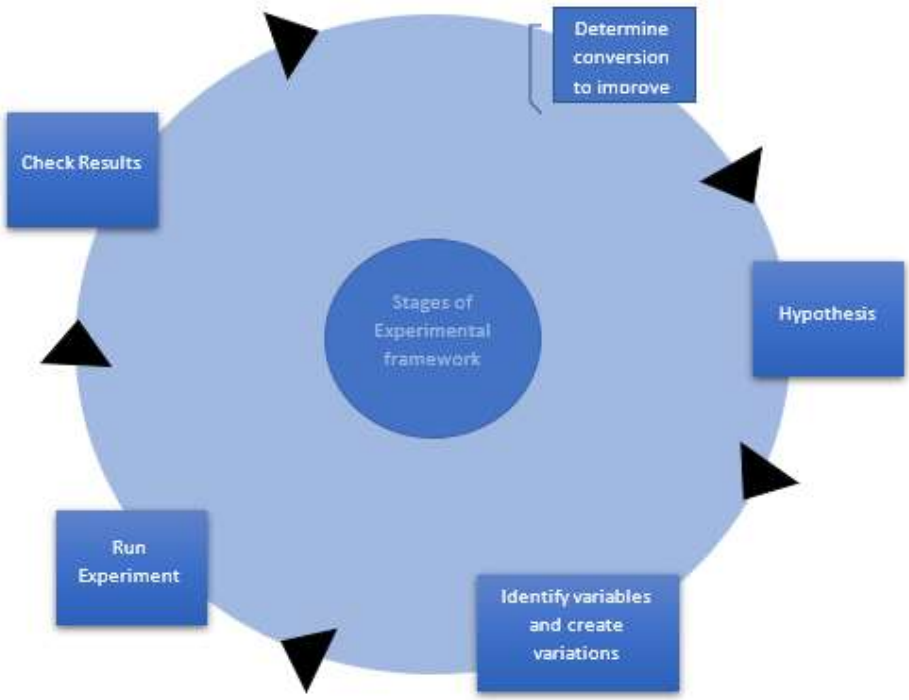


Fig 5.3 Procedure to perform test

5.4 STATISTICAL SIGNIFICANCE OF THE TEST

If we want to reject our null hypothesis, we must first establish that the outcomes of our tests are statistically significant. Our hypothesis testing may be hampered by inaccuracies that manifest themselves in one of the following ways:

We make a Type I mistake when we reject the null hypothesis despite the fact that it is accurate. In other words, we will accept version B if it does not surpass variation A in terms of performance and is thus superior to it.
A type II error occurred when we failed to reject the null hypothesis after it was shown to be false. So, when version B outperforms variation A, we infer that version B is not outstanding. We must calculate the statistical significance of our test in order to avoid making these mistakes.

We consider an experiment to be statistically significant when there is sufficient evidence to demonstrate that the result observed in the sample is also observed in the population as a whole.

This signifies that the difference between your control version and the test version is not the result of a mistake or a random chance event. A two-sample T-test can be used to determine whether or not our experiment was statistically significant.

It is one of the most often used hypothesis tests, and it is a two–sample t–test. It is used to determine whether the average difference between the two groups is greater than zero.

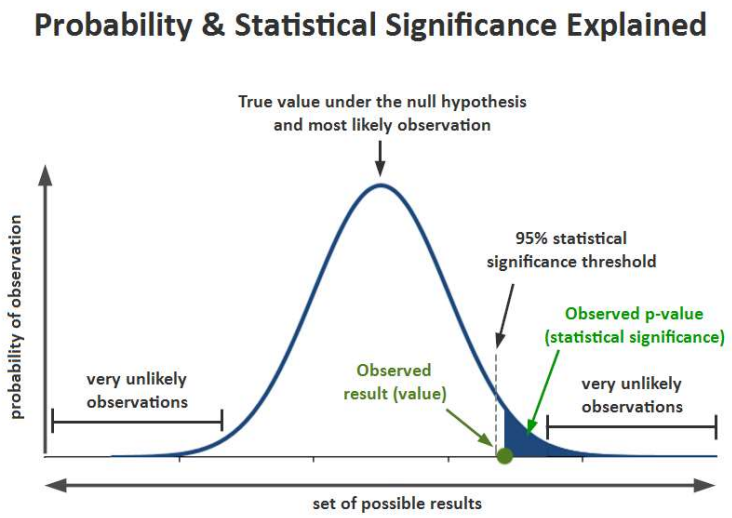


Fig 5.4 Probability and Statistical significance

It is the chance of rejecting a null hypothesis when the null hypothesis is true that is measured by the significance level in statistics (also known as alpha or). Generally, we use a 0.05 threshold of significance for our statistical analyses.

P-Value: When a disparity between two numbers is the consequence of pure chance, this term

is used to describe the probability that the discrepancy occurred by chance. The p-value suggests that there is evidence that the null hypothesis is incorrect. When the p-value is lower, there is a larger possibility of rejecting the null hypothesis. For a significance threshold of 0.05, we can rule out the null hypothesis since it is less significant than the null hypothesis with a p-value of less than 0.05.

The confidence interval is as follows: When a particular percentage of test results fall inside a given range, the confidence interval is defined as follows: At the beginning of our test, we manually select the confidence level that we want to use. In most cases, we use a 95 percent confidence interval for our calculations. The sample graph and how to read it is explained in Figure 5.4. The formula for T statistics is

$$T - statistic = \frac{Observed\ value - hypothesized\ value}{Standard\ Error}$$

$$Stamdard\ Error = \sqrt{\frac{2 * Variance(sample)}{N}}$$

CHAPTER 6

OUTPUT AND EXPLANATION

OUTPUT

With this thesis, the scope of A/B testing, which is a type of online controlled experiment, was broadened from being a tool that is often used to evaluate the aesthetic components of websites to a tool that can be used in E-commerce to assess the impact of variations in sales processes, which was the overall goal. The Decision Assistant tool, which featured native A/B testing capabilities and was designed specifically for the purpose of developing variations of a sales process, was explored in detail as part of this thesis as well. A series of continuing meetings with colleagues enabled us to get valuable insight into what constitutes a successful online sales process, as well as where A/B testing should be emphasized. Following the aforementioned procedure, hypotheses were developed and tested online using A/B test trials, which were generated as a consequence of the method described above. However, further study may be necessary to confirm whether or not it is practical to increase A/B testing in this area. As a consequence of the study, a series of suggestions was established to guarantee that such a feat might be effectively accomplished by enterprises operating in the sector. By including these ideas, as well as any new criteria that may be discovered via further research, it is feasible to construct a general framework for utilizing A/B testing as a tool for evaluating sales processes.

EXPLANATION

The purpose of this project was to explain how a business may motivate its workers to run A/B testing in a manner that adds to long-term company-wide optimization — and we offered a rudimentary framework for how the company could communicate data with the experimenters in order to do this. Even in the absence of private information, this Minimum Learning Mechanism perfectly aligns incentives, and it remains resilient even when workers utilize a little amount of private information to alter the designs of their trials.

The MLM can be used in a variety of situations, including but not limited to:

- In terms of the company's structure, we have imposed no restrictions other than the As

for the organizational structure, we have set no limits beyond the condition that the treatment effect in an experiment may be depicted at any moment as an IID draw from a distribution indexed by certain factors that are common to all workers.

- For each trial, we set only mild limits on the distributions of treatment effects and observations, and we place no restrictions on the distributions of observations. In this case, we simply assume that calculating the posterior for the impact is tractable, making a Bayesian analysis viable.
- We are agnostic as to why the experimenters want to draw conclusions about treatment effects in the first place, and we are not interested in finding out. The following assumptions will be made to keep things simple: the goal is to maximize the reward associated with a decision that can be made instantly, that the reward associated with every consequence under any option is fully understood, and that maximizing the posterior expected reward is computationally feasible.

CHAPTER 7

CONCLUSION AND FUTURE WORK

CONCLUSION

As a result, data-driven choices have the greatest influence on production and productivity when it comes to operating a company, which is why they are so important. Companies perform experiments on a regular basis in order to aid in decision-making processes. Throughout this article, we demonstrate how systematic evaluation is carried out in the field of software engineering, as well as how the results are analyzed after they have been collected. The capacity of clients' software to connect with one another through the internet has increased the breadth of online experimentation and has made it more accessible to more people. As a consequence, experimentation becomes much more scalable and efficient, and findings can be evaluated much more quickly as a result of this. It is necessary to use the Bayesian approach in order to make interpretation easier. Also included was an explanation of the three most common assessment situations as well as a solution for one of them. Continuous trials and the A/B test, we suggest, are potentially lucrative research topics since there are several research gaps to be addressed in this domain.

FUTURE WORK

Throughout this thesis, we have discussed four of the most pressing issues facing A/B testing practitioners today, which we feel to be the most critical challenges. These are the ones to look out for:

1. Monitoring on a constant basis is essential.
2. Temporal acoustic noise
3. The use of a large number of inferences.
4. The region is underexplored at the moment.

To address each of the four challenges, we have developed distinct solutions that are easy enough for an online organization to incorporate into its A/B testing infrastructure with no extra work. It is possible to calculate p-values and confidence intervals using this technique, and the results may be conveyed to experimenters using the same interface as the standard t-testing

approach. This method of computing p-values and confidence intervals is an alternative to the traditional method.

The authors offer a basic technique to combating under-exploration that enables the firm to select which data from past A/B experiments each experimenter should be authorized to access. When it comes to statistical trade-offs, such as power versus run-time, one of the most significant advantages of our methodology is that it is completely independent of the types of decisions that are being optimized through A/B testing as well as the specific preferences of each experimenter when it comes to power versus run-time. When working with big, decentralized organizations, where the use cases for A/B testing are exceedingly different, this level of flexibility becomes crucial.

However, as we have proven in this study, there is still much more work to be done, both to enhance our responses to the four issues we have identified and to help in the design and analysis of A/B testing in general, as we have demonstrated. In this part, we will outline some potential study opportunities for the foreseeable future.

According to the objectives of this thesis, the A/B testing approach for business improvement was used, with the quality of every new concept being summarised by a difference in a mean response per visitor under the treatment and control circumstances, respectively. For statistical conclusions to be made about this difference, it is essential to collect further data, and the experimenter may use the statistical inference to direct her toward whatever option she feels would enhance service. Using an A/B testing viewpoint has the benefit of allowing researchers to reach conclusions on their own merits rather than setting rigorous restrictions on the sorts of possibilities they are permitted to choose from.

This has the drawback of not being directly tailored to any given decision-making goal since the statistical conclusions that are created are not. This is due to a restriction of the technique.

Approaches to business optimization that are based on reinforcement learning, on the other hand, are gaining in popularity. Before they can begin optimising choices when new information becomes available, they must first specify the set of possible options and the incentives associated with any outcomes that are to be considered. In industries such as targeted advertising and viral marketing, where the goals are clear and little dynamic human interaction is required, these types of algorithms are extremely popular.

REFRENCES

[1] Shafi Kamalbasha and Manuel J. A. Eugster Avira Operations GmBH & Co. KG Tettnang, Germany. Bayesian A/B Testing for Business Decisions. https://doi.org/10.48550/arXiv.2003.02769.

[2] Giordano Tamburrelli and Alessandro Margara Faculty of Informatics. University of Lugano, Switzerland. {giordano. Tamburrelli alessandro.margara}@usi.ch Towards Automated A/B Testing. https://www.researchgate.net/publication/264435539

[3] Andrea Arcuri and Lionel Briand Simula Research Laboratory, P.O. Box 134, Lysaker, Norway. Email: {arcuri,briand}@simula.no. A Hitchhiker's Guide to Statistical Tests for Assessing Randomized Algorithms in Software Engineering. Technical Report, Simula Research Laboratory, no. 2011-13.

[4] Frank P.A. Coolen. Dept. of Mathematical Sciences, Durham University, Durham, United Kingdom frank.coolen@durham.ac.uk. Some statistical aspects of software testing and reliability. https://maths.durham.ac.uk/stats/people/fc/Springer 12-preprint2.pdf

[5] Ron Kohavi1 and Roger Longbotham2 1Application Services Group, Microsoft, Bellevue, WA, USA 2Microsoft, Data and Decision Sciences Group, Redmond, WA, USA. Online Controlled Experiments and A/B Testing. DOI: 10.1007/978-1- 4899-7687-1_891. https://www.researchgate.net/publication/316116834_Online_Controlled_Experiments_and_AB_Testi ng.

[6] Rasmus Ros rasmus.ros@cs.lth.se Lund University Sweden. Per Runeson per.runeson@cs.lth.se Lund University Sweden. Continuous Experimentation and A/B Testing: A Mapping Study. 2018 ACM/IEEE 4th International Workshop on Rapid Continuous Software Engineering. https://ieeexplore.ieee.org/document/8452106

[7] Nikhil Bhat, Graduate School of Business Columbia University. Vivek F. Farias, Sloan School of Management Massachusetts Institute of Technology. Camac C. Moallemi, Graduate School of Business Columbia University. Deeksha Sinha Operations Research Center Massachusetts Institute of Technology.

https://web.mit.edu/vivekf/www/papers/experiment_design.pdf.

[8] Maria Esteller-Cucala, Universitat Politècnica de Catalunya Barcelona Tech., Barcelona, Spain. Vicenc Fernandez, SEAT, S.A., Barcelona, Spain. Diego Villuendas TechTalent-Lab, Department of Management, Universitat Politècnica de CatalunyaBarcelona Tech., Barcelona, Spain. The AB Testing Pitfalls Companies Might Not Be Aware of—A Spotlight on the Automotive Sector Websites. https://doi.org/10.3389/frai.2020.00020

[9] Amy Gallo, A Refresher on A/B Testing. Harvard Business Review. Analytics and Data Science. https://hbr.org/2017/06/a-refresher-on-ab-testing

[10] Chopra Paras, " A/B Testing Guide", "Everything you need to know about A/B Testing - from why to how, challenges to examples", https://vwo.com/ab-testing/

[11] Simpson Jon, "A/B Testing: The Benefits And How To Use It Efficiently", https://www.forbes.com/sites/forbesagencycouncil/2020/03/12/ab-testing-the-benefits-and-how-to-use-it efficiently/?sh=e90800e786d4

[12] Patel Nell, "A/B Testing: Definition, How it Works, Examples, & Tools", https://neilpatel.com/blog/ab-testing-introduction/

[13] Komang Candra Brata, Adam Hendra Brata Mobile, Game, and Media Research Group, Department of Computer Science, Universitas Brawijaya, Indonesia. User experience improvement of Japanese language mobile learning application through the mental model and A/B testing. DOI: 10.11591/ijece.v10i3.pp2659-2667

[14] Isak Kabir, How to conduct A/B Testing? https://towardsdatascience.com/how-to-conduct-a-b-testing-3076074a8458

[15] A. Gelman, J. Carlin, H. Stern, D. Dunson, A. Vehtari, and D. Rubin, Bayesian Data Analysis, 3rd ed. Chapman and Hall/CRC, 2013

[16] J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[17] I. S. Jacobs and C. P. Bean, "Fine particles, thin films and exchange anisotropy," in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New

York: Academic, 1963, pp. 271–350.

[18] K. Elissa, “Title of paper if known,” unpublished. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[19] Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].

[20] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

APPENDIX 1

SAMPLE CODE

Here is the sample code that we used to perform Bayesian test

```
create or replace temp view testing_data_ar_opt as
select allocation_dt,
count(distinct case when cohort_name = 'a' then client_id else null end) as total_devices_control,
count(distinct case when status_as_of_allocation = 'NAR' and status_as_of_run_date = 'AR' and cohort_name = 'a' then client_id else null end) as opt_in_control,
count(distinct case when cohort_name = 'b' then client_id else null end) as total_devices_test,
count(distinct case when status_as_of_allocation = 'NAR' and status_as_of_run_date = 'AR' and cohort_name = 'b' then client_id else null end) as opt_in_test
from bi_workspace.pos_ar_opt_in_launch_2_results
where date(allocation_dt)>='2022-02-15'
GROUP by 1
order by 1

OK
```

```
%py
df2 = spark.sql("""select * from testing_data_ar_opt""").toPandas()
df2.shape

Out[2]: (11, 5)
```

Fig A 1.1 Sample Code 1

```
%py
test_total = df2['total_devices_test'].tolist()
test_opt_in = df2['opt_in_test'].tolist()
control_total =df2['total_devices_control'].tolist()
control_opt_in = df2['opt_in_control'].tolist()
```

```
%py
import pandas as pd
import numpy as np
# setting seed to reproduce the results
np.random.seed(0)

control_conversions = []
test_conversions = []
bayesian_p_value = []

bayesian_loss_test = []
bayesian_loss_control = []

bayesian_lower_diff = []
bayesian_upper_diff = []
```

Fig A 1.2 Sample Code 2

```
%py
from scipy.stats import norm, beta
from statsmodels.stats.proportion import proportions_ztest
import matplotlib.pyplot as plt
import seaborn as sns
sns.set()
%matplotlib inline
# Prior: Taking uniform Prior for bayesian
a = b = 1

# Number of samples for sampling
N = 100000

# Starting at day 1 as then is when we have significant population for testing
start = 1

for i in range(start, len(control_total) + 1):

    # Creating cumulative as of date status for conversions
    control_c = sum(control_opt_in[:i])
    control_t = sum(control_total[:i])
    test_c = sum(test_opt_in[:i])
    test_t = sum(test_total[:i])

    # Calculating conversion rates for both variants
   # control_conversions.append(control_c/control_t)
   # test_conversions.append(test_c/test_t)
```

Fig A 1.3 Sample Code 3

```
  # Posterior distribution for Bayesian
  control_dist = beta(control_c + a, control_t - control_c + b)
  test_dist = beta(test_c + a, test_t - test_c + b)

 # Sampling from posterior distribution
  control_samples = control_dist.rvs(N)
  test_samples = test_dist.rvs(N)

  # Lift distribution
  diff_dist = test_samples - control_samples

  # 95% HPD interval for Lift
  bayesian_lower_diff.append(np.percentile(diff_dist, 2.5))
  bayesian_upper_diff.append(np.percentile(diff_dist, 97.5))

  # Bayesian p-value - Test better than Control
  bayesian_prob = sum(test_samples > control_samples)/N

  treatment_won = [i <= j for i,j in zip(control_samples, test_samples)]

  # Expected Test and Control Loss
  loss_control = [max(j - i, 0) for i,j in zip(control_samples, test_samples)]
  loss_treatment = [max(i - j, 0) for i,j in zip(control_samples, test_samples)]

  all_loss_control = [int(i)*j for i,j in zip(treatment_won, loss_control)]
  all_loss_treatment = [(1 - int(i))*j for i,j in zip(treatment_won, loss_treatment)]
```

Fig A 1.4 Sample Code 4

```
all_loss_control = [int(i)*j for i,j in zip(treatment_won, loss_control)]
all_loss_treatment = [(1 - int(i))*j for i,j in zip(treatment_won, loss_treatment)]

expected_loss_control = np.mean(all_loss_control)
expected_loss_treatment = np.mean(all_loss_treatment)

bayesian_loss_control.append(expected_loss_control)
bayesian_loss_test.append(expected_loss_treatment)

bayesian_p_value.append(bayesian_prob)
```

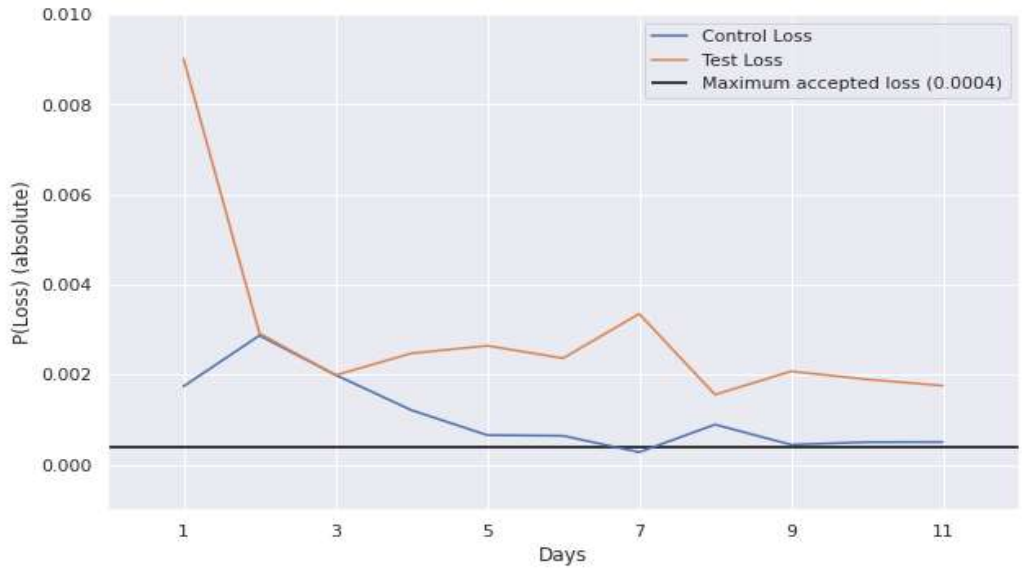
```
%py
# Absolute Loss
loss_threshold = 0.0004
day_start = 1

fig=plt.figure(figsize=(10,6))
plt.plot(range(day_start, len(bayesian_loss_control) + day_start), bayesian_loss_control, label='Control Loss')
plt.plot(range(day_start, len(bayesian_loss_control) + day_start), bayesian_loss_test, label='Test Loss')
plt.hlines(loss_threshold, day_start-1, len(bayesian_loss_control) + day_start, color='black', label=f'Maximum accepted loss ({loss_threshold})')
plt.ylim([-0.001,0.01])
plt.xticks(range(day_start, len(bayesian_loss_control) + day_start, 2))
plt.xlim(day_start-1, len(bayesian_loss_control) + day_start)
plt.ylabel('P(Loss) (absolute)')
plt.xlabel('Days')
```

Fig A 1.5 Sample Code 5

```
plt.ylabel('P(Loss) (absolute)')
plt.xlabel('Days')
plt.legend(loc='upper right')
```

```
Out[7]:
```




```
<matplotlib.legend.Legend at 0x7fb6163c5940>
```

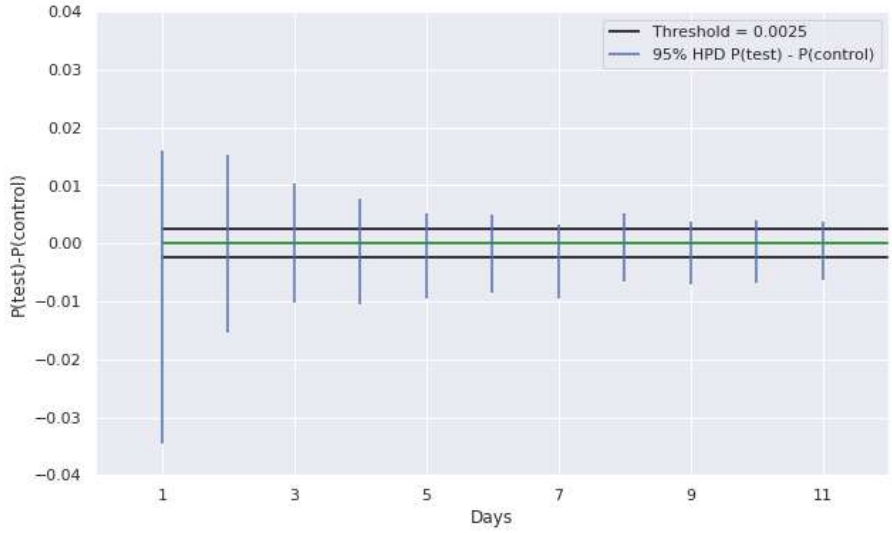
Fig A 1.6 Sample Code 6

```
# Absolute Loss
loss_threshold = 0.0025
day_start = 1

fig=plt.figure(figsize=(10,6))
plt.hlines(0, day_start, len(bayesian_loss_control) + day_start, color='green')
plt.hlines(loss_threshold, day_start, len(bayesian_loss_control) + day_start, color='black', label=f'Threshold = {loss_threshold}')
plt.hlines(-loss_threshold, day_start, len(bayesian_loss_control) + day_start, color='black')
for i in range(len(bayesian_lower_diff)):
    if not i:
        plt.vlines(i+day_start, bayesian_lower_diff[i], bayesian_upper_diff[i], label=f'95% HPD P(test) - P(control)')
    else:
        plt.vlines(i+day_start, bayesian_lower_diff[i], bayesian_upper_diff[i])
plt.ylim(-0.04, 0.04)
plt.ylabel('P(test)-P(control)')
plt.xlabel('Days')
plt.xticks(range(day_start, len(bayesian_loss_control) + day_start, 2))
plt.xlim(day_start-1, len(bayesian_loss_control) + day_start)
plt.legend()

Out[8]:
```

Fig A 1.7 Sample Code 7




```
<matplotlib.legend.Legend at 0x7fb616358f10>
```

```
%py
print(bayesian_p_value)

[0.24785, 0.49723, 0.49999, 0.38628, 0.29081, 0.30179, 0.16345, 0.41193, 0.27005, 0.29814, 0.31075]
```

Fig A 1.8 Sample Code 8

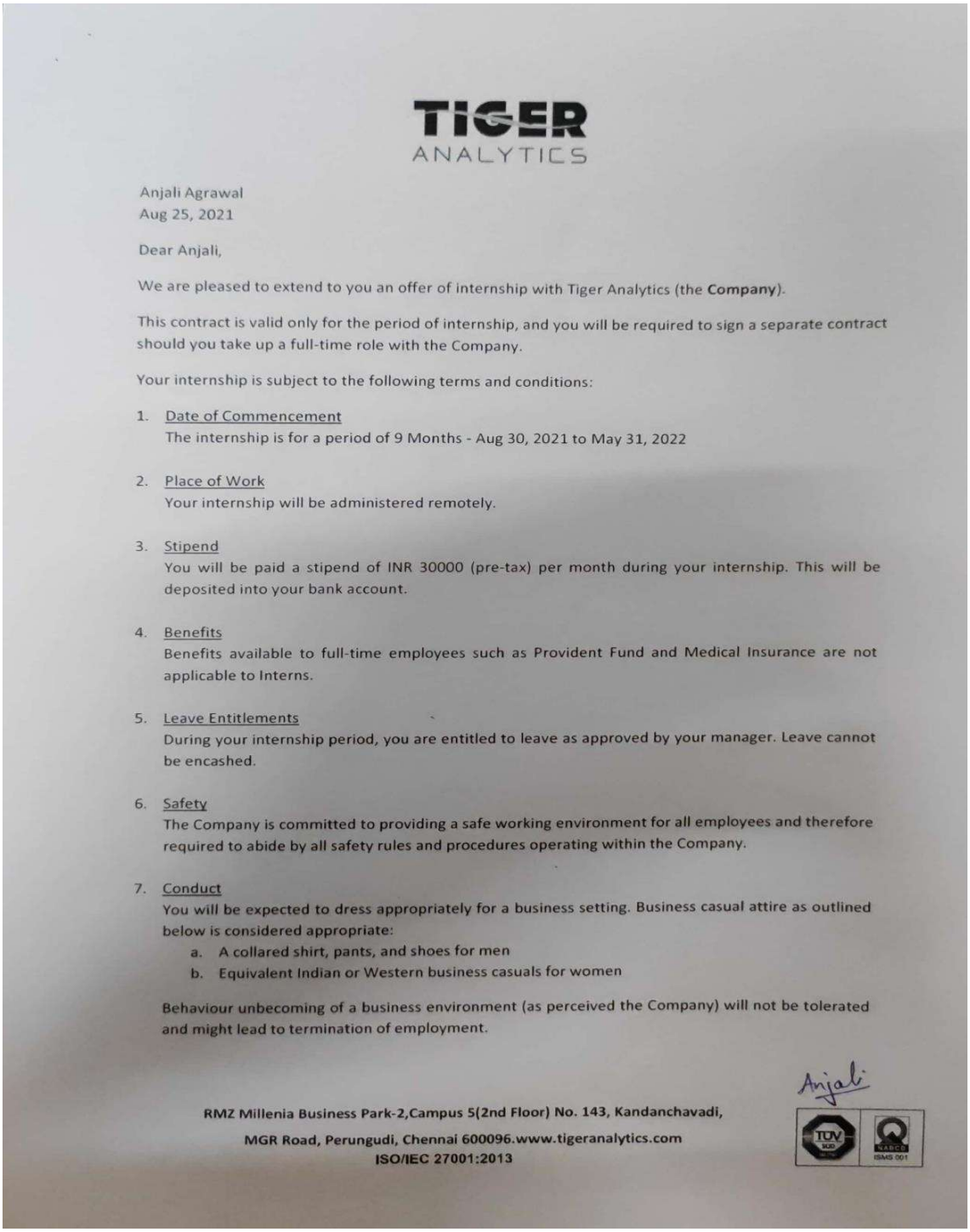
```
%py
print(bayesian_prob)

0.31075
```

Fig A 1.9 Sample Code 9

APPENDIX 2

INTERNSHIP LETTER



TIGER ANALYTICS

Anjali Agrawal
Aug 25, 2021

Dear Anjali,

We are pleased to extend to you an offer of internship with Tiger Analytics (the **Company**).

This contract is valid only for the period of internship, and you will be required to sign a separate contract should you take up a full-time role with the Company.

Your internship is subject to the following terms and conditions:

1. Date of Commencement
   The internship is for a period of 9 Months - Aug 30, 2021 to May 31, 2022

2. Place of Work
   Your internship will be administered remotely.

3. Stipend
   You will be paid a stipend of INR 30000 (pre-tax) per month during your internship. This will be deposited into your bank account.

4. Benefits
   Benefits available to full-time employees such as Provident Fund and Medical Insurance are not applicable to Interns.

5. Leave Entitlements
   During your internship period, you are entitled to leave as approved by your manager. Leave cannot be encashed.

6. Safety
   The Company is committed to providing a safe working environment for all employees and therefore required to abide by all safety rules and procedures operating within the Company.

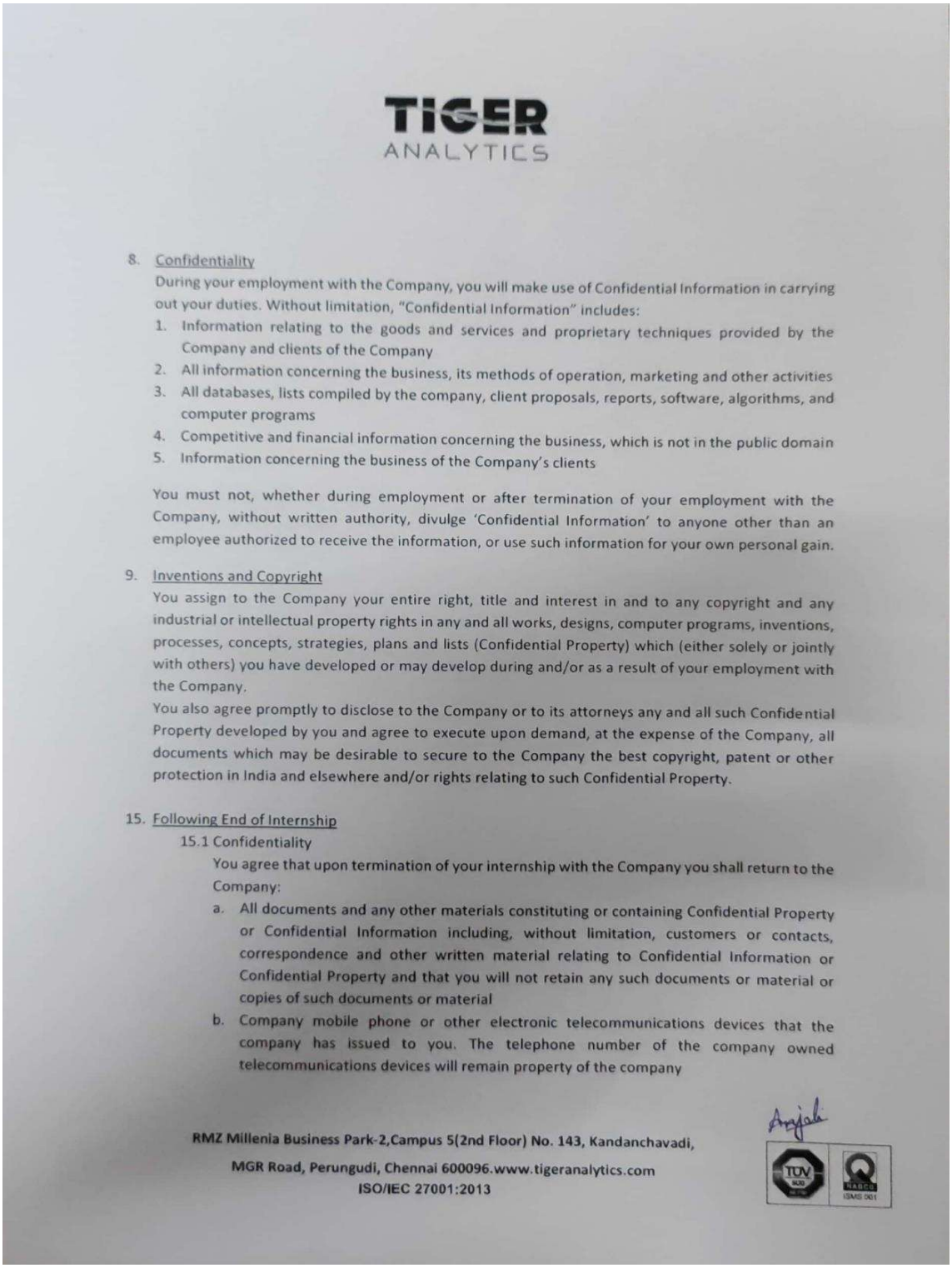
7. Conduct
   You will be expected to dress appropriately for a business setting. Business casual attire as outlined below is considered appropriate:
   a. A collared shirt, pants, and shoes for men
   b. Equivalent Indian or Western business casuals for women

Behaviour unbecoming of a business environment (as perceived the Company) will not be tolerated and might lead to termination of employment.

Anjali

RMZ Millenia Business Park-2,Campus 5(2nd Floor) No. 143, Kandanchavadi,
MGR Road, Perungudi, Chennai 600096.www.tigeranalytics.com
ISO/IEC 27001:2013

Fig A 2.1 Offer Letter p1



TIGER ANALYTICS

8. Confidentiality

During your employment with the Company, you will make use of Confidential Information in carrying out your duties. Without limitation, "Confidential Information" includes:

1. Information relating to the goods and services and proprietary techniques provided by the Company and clients of the Company
2. All information concerning the business, its methods of operation, marketing and other activities
3. All databases, lists compiled by the company, client proposals, reports, software, algorithms, and computer programs
4. Competitive and financial information concerning the business, which is not in the public domain
5. Information concerning the business of the Company's clients

You must not, whether during employment or after termination of your employment with the Company, without written authority, divulge 'Confidential Information' to anyone other than an employee authorized to receive the information, or use such information for your own personal gain.

9. Inventions and Copyright

You assign to the Company your entire right, title and interest in and to any copyright and any industrial or intellectual property rights in any and all works, designs, computer programs, inventions, processes, concepts, strategies, plans and lists (Confidential Property) which (either solely or jointly with others) you have developed or may develop during and/or as a result of your employment with the Company.

You also agree promptly to disclose to the Company or to its attorneys any and all such Confidential Property developed by you and agree to execute upon demand, at the expense of the Company, all documents which may be desirable to secure to the Company the best copyright, patent or other protection in India and elsewhere and/or rights relating to such Confidential Property.

15. Following End of Internship

15.1 Confidentiality

You agree that upon termination of your internship with the Company you shall return to the Company:

a. All documents and any other materials constituting or containing Confidential Property or Confidential Information including, without limitation, customers or contacts, correspondence and other written material relating to Confidential Information or Confidential Property and that you will not retain any such documents or material or copies of such documents or material
b. Company mobile phone or other electronic telecommunications devices that the company has issued to you. The telephone number of the company owned telecommunications devices will remain property of the company

Anjali

Fig A 2.2 Offer Letter p2

TIGER
ANALYTICS

c. And other property of the Company including, without limitation, security access cards, credit cards, computers and computer software (which must be deleted immediately from any storage device owned by you).

15.2 Non-rendering of services to clients

You also agree that you will not, during your employment with the Company or Group and for a period of twelve months after termination, contact, or assist anyone else to contact, any Client of the Company with a view to obtaining business from those Clients; or to persuade those Clients to cease to do business with the Company or to reduce the amount of business which any such Clients has customarily done or is reasonable expected to do with the Company.

The definition of Client includes any prospective client to whom you or the Company has made a formal presentation at any time during the twelve months immediately preceding termination.

You also agree that, if such clients approach you in that period, you will inform them of this agreement and refer them to the appropriate person at the Company.

15.3 Non-hire of staff

You also agree that, for six months following termination of your employment with the Company, you will not hire or encourage a third party to hire any member of staff of the Company.

16. Invalidity

In any terms of provisions in this agreement shall be held illegal or unenforceable, in whole or in part, under any enactment or rule of law, such term or provision or part shall to that extent be deemed not to form part of this agreement but the enforceability of the remainder of this agreement shall not be affected.

17. Variation

The terms of this contract of employment may be varied by the Company from time to time. You will be notified of any variations.

18. Adherence to Company Policies

When you join the Company, it will also be a condition of employment that you review and adhere to company policies which you will be notified of subsequently. You agree to adhere to the Company's project financing contracts (e.g. BOT) with the clients.

19. Governing Laws and Jurisdiction

This contract will be governed by the law in force in Chennai, India.

Fig A 2.3 Offer Letter p3

TIGER
ANALYTICS

20. Personal Information and Consent

By accepting this offer, you are giving your implicit consent to Tiger Analytics to collect and use your personal information for business purposes. Your personal information may be shared with the Clients and prospective Clients of Tiger Analytics as a part of selection or onboarding process to work in projects. Tiger Analytics will also share your personal information with a third party for carrying out the background verification as required. Tiger Analytics will store your employment, financial and personal information during the period of employment and for Data Retention Period after your separation, as per the data retention policy to comply with statutory requirements.

21. Acceptance

Please sign this letter signifying your acceptance of the appointment and the conditions of service specified in this letter.

We are pleased to welcome you to the Company. If the preceding terms and conditions of your employment with the Company are acceptable to you, please indicate you acceptance by initialing each page and signing the last page of the attached copy and returning it to me.

Regards

G. Pradeep kumar

Pradeep Gulipalli
General Manager
Tiger Analytics India LLP

Fig A 2.4 Offer Letter p4

APPENDIX 3

TECHNICAL PAPER PRESENTATION

My link to the conference paper is: https://tinyurl.com/ab-test-paper-by-anjali

I have presented in the 10th International Conference on Contemporary Engineering and 1technology 2022 (ICCET2022).

10TH INTERNATIONAL CONFERENCE ON CONTEMPORARY ENGINEERING AND TECHNOLOGY 2022

ORGANIZED BY

ORGANIZATION OF SCIENCE & INNOVATIVE ENGINEERING AND TECHNOLOGY (OSIET), CHENNAI, INDIA.

IN ASSOCIATION WITH

PRINCE SHRI VENKATESHWARA PADMAVATHY ENGINEERING COLLEGE

PRINCE DR. K. VASUDEVAN COLLEGE OF ENGINEERING & TECHNOLOGY

(Approved By All India Council For Technical Education, Affiliated To Anna University)

Medavakkam - Mambakkam Main Road, Ponmar, Chennai - 600 127.

Certificate of Registration

This is to certify that Anjali Agrawal from SRM Institute of Science and Technology, Kattankulathur has submitted a paper titled A/B Testing for Industrial comapanies has registered in the "10th International Conference on Contemporary Engineering and Technology 2022" will be held on 30th April & 1st May 2022 and invited for the Presentation

Dr. P. Aravinthan, M.E., Ph.D., Co-Convener - ICCET

Dr. A. Krishnamoorthy, M.E., Ph.D Convener - ICCET

K.Janani, M.Tech., CEO, OSIET

Fig A 3.1 Certificate 1

Fig A 3.2 Certificate 2

Fig A 3.3 Certificate 3

PLAGIARISM REPORT

Anjaliv2

ORIGINALITY REPORT

8%	6%	2%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	splitmetrics.com Internet Source	3%
2	vwo.com Internet Source	2%
3	Frank P. A. Coolen. "Chapter 7 On Some Statistical Aspects of Software Testing and Reliability", Springer Science and Business Media LLC, 2013 Publication	1%
4	Florian Auer, Rasmus Ros, Lukas Kaltenbrunner, Per Runeson, Michael Felderer. "Controlled experimentation in continuous experimentation: Knowledge and challenges", Information and Software Technology, 2021 Publication	<1%
5	appliedmarketinggroup.biz Internet Source	<1%
6	arxiv.org Internet Source	<1%

7	uxdesign.cc Internet Source	<1%
8	onlinelibrary.wiley.com Internet Source	<1%
9	lucris.lub.lu.se Internet Source	<1%
10	Blatchley, Barbara. "Statistics in Context", Oxford University Press Publication	<1%
11	Submitted to Kaplan Professional Student Paper	<1%
12	Submitted to University of Liverpool Student Paper	<1%
13	Submitted to University of North Texas Student Paper	<1%
14	Submitted to American Public University System Student Paper	<1%
15	Submitted to Technological Institute of the Philippines Student Paper	<1%
16	Submitted to The Hong Kong Polytechnic University Student Paper	<1%
17	cdr.lib.unc.edu	

	Internet Source	<1%
18	Submitted to Loughborough University Student Paper	<1%
19	"Proven, Practical Guidance from The Planned Giving Experts", Planned Giving Today, 2016 Publication	<1%
20	Dileeka Dias. "Logic analysers", Institution of Engineering and Technology (IET), 2003 Publication	<1%
21	Submitted to Southern New Hampshire University - Continuing Education Student Paper	<1%

Exclude quotes	On	Exclude matches	< 10 words
Exclude bibliography	On		

Format - I

SRM INSTITUTE OF SCIENCE AND TECHNOLOGY

(Deemed to be University u/s 3 of UGC Act, 1956)

Office of Controller of Examinations

REPORT FOR PLAGIARISM CHECK ON THE DISSERTATION/PROJECT REPORTS FOR UG/PG PROGRAMMES

(To be attached in the dissertation/ project report)

1	Name of the Candidate **(IN BLOCK LETTERS)**	ANJALI AGRAWAL
2	Address of the Candidate	Ashok Medical Store, Dharamsala Road Padrauna, 274304, U.P. **Mobile Number :** 8939351124
3	Registration Number	RA1811030010096
4	Date of Birth	28/04/2000
5	Department	Computer Science and Engineering (CS)
6	Faculty	Faculty of Enginnering & Technology
7	Title of the Dissertation/Project	A/B Testing for Industrial companies
8	Whether the above project/dissertation is done by	Individual or [redacted] : (Strike whichever is not applicable) a) If the project/ dissertation is done in group, then how many students together completed the project : 1 b) Mention the Name & Register number of other candidates : Anjali Agrawal: RA1811030010096
9	Name and address of the Supervisor / Guide	Dr. S. Prabakeran, Assistant Professor Department of Networking and Communications SRM Institute of Science & Technology +91 9042394880 prabakes@srmist.edu.in **Mail ID : Mobile Number :**
10	Name and address of the Co-Supervisor / Co- Guide (if any)	**Mail ID : Mobile Number :**

11	Software Used	ouriginal , Quillbot		
12	Date of Verification	25/04/2022		
13	**Plagiarism Details: (to attach the final report from the software)**			
Chapter	**Title of the Chapter**	**Percentage of similarity index (including self citation)**	**Percentage of similarity index (Excluding self citation)**	**% of plagiarism after excluding Quotes, Bibliography, etc.,**
1	Introduction	5.2%	3.3%	<1%
2	Literature Review	2.2%	1.4%	<1%
3	System Analysis	1.3%	0.8%	<1%
4	System Design	5.1%	4.1%	<1%
5	System Implementation	6.3%	5.9%	1.2%
6	Output and Explanation	0%	0%	<1%
7	Conclusion and Future Work	3.1%	0.6%	<1%
8				
9				
10				
Appendices		8%	7.1%	4.2%

I / We declare that the above information have been verified and found true to the best of my / our knowledge.

Signature of the Candidate

Name & Signature of the Staff (Who uses the plagiarism check software)

Name & Signature of the Supervisor/Guide

Name & Signature of the Co-Supervisor/Co-Guide

Name & Signature of the HOD

PAPER ACCEPTANCE PROOF

Attached below is the publication proof provided by ICCET and EASR for my research paper "A/B testing for industrial companies".

Engineering and Applied Science Research (EASR)

ISSN 2539-6161| ISSN 2539-6218

Dear Author/s, 30th April 2022

Thank you for your submission and for your interest in the Engineering and Applied Science Research. The below mentioned article/s has been successfully passed the initial review process. The article/s is now sent for further review process to ensure that all publication requirements have been met and it is expected to get confirmed by Jun/Jul 2022 unless otherwise specified.

PAPER TITLE	AUTHOR1	AUTHOR2	AUTHOR3	AUTHOR4
MESH MESSAGING SERVICE USING BLUETOOTH	Aswini.J	P. Hitesh Reddy	R.B.Moulya	P. Ruchitha, P. Ashish & S.Amisha
IMPROVING THE LIQUID INSULATION PERFORMANCE WITH ANTIOXIDANTS	Helan Gethse J	Dhanya K	Diana Hyden N	Muthuselvi G, Samuel Pakianathan P
ALCOHOL DETECTION AND EMERGENCY ALERT SYSTEM	PREM RAJA R	BARKAVI G	AISHWARYA S	KEERTHANA K, VASUDEVAN V
A/B TESTING FOR INDUSTRIAL COMAPANIES	Anjali Agrawal	Dr. S. Prabakeran	Preeti Vijayan	
IMPROVED DIGITAL TWINNING TECHNIQUES FOR SMART SIMULATIONS	Dr.SHAIK JAVED PARVEZ	AKASH VENKATESAN	VANRAJ D	

Thank you again for your submission.

Regards,
Editor