OPTIMIZING AI-ASSISTED LOAN APPROVALS

TABLE OF CONTENTS

INTRODUCTION

It's a fact that in domains where money is at stake, a human's cognitive constraints and behavioural biases may impede the processing and interpreting of less salient, non-quantitative information effectively and undermine their decision-making (Campbell, et al., 2019). These biases result in suboptimal decisions on loan approvals and increase financial risk for the lending institutions (Oyeniyi, et al., 2022). To optimally allocate resources and check the credibility of the model, institutions often leverage artificial intelligence (AI) to assess whether a newly developed machine learning model can enhance the decision-making process for loan officers in a consumer lending company (Amato, et al., 2024).

Assumptions:

The findings of this report have the following assumptions:

1. Loan officers follow a two-stage review process: an initial independent assessment (1) and a final review with AI assistance (2).
2. Officers are randomly allocated to treatment and control groups.
3. There are no systematic biases between groups before and after AI's decision.
4. The error rates (Type I & Type II) are measured the same way across groups.

Company Objective:

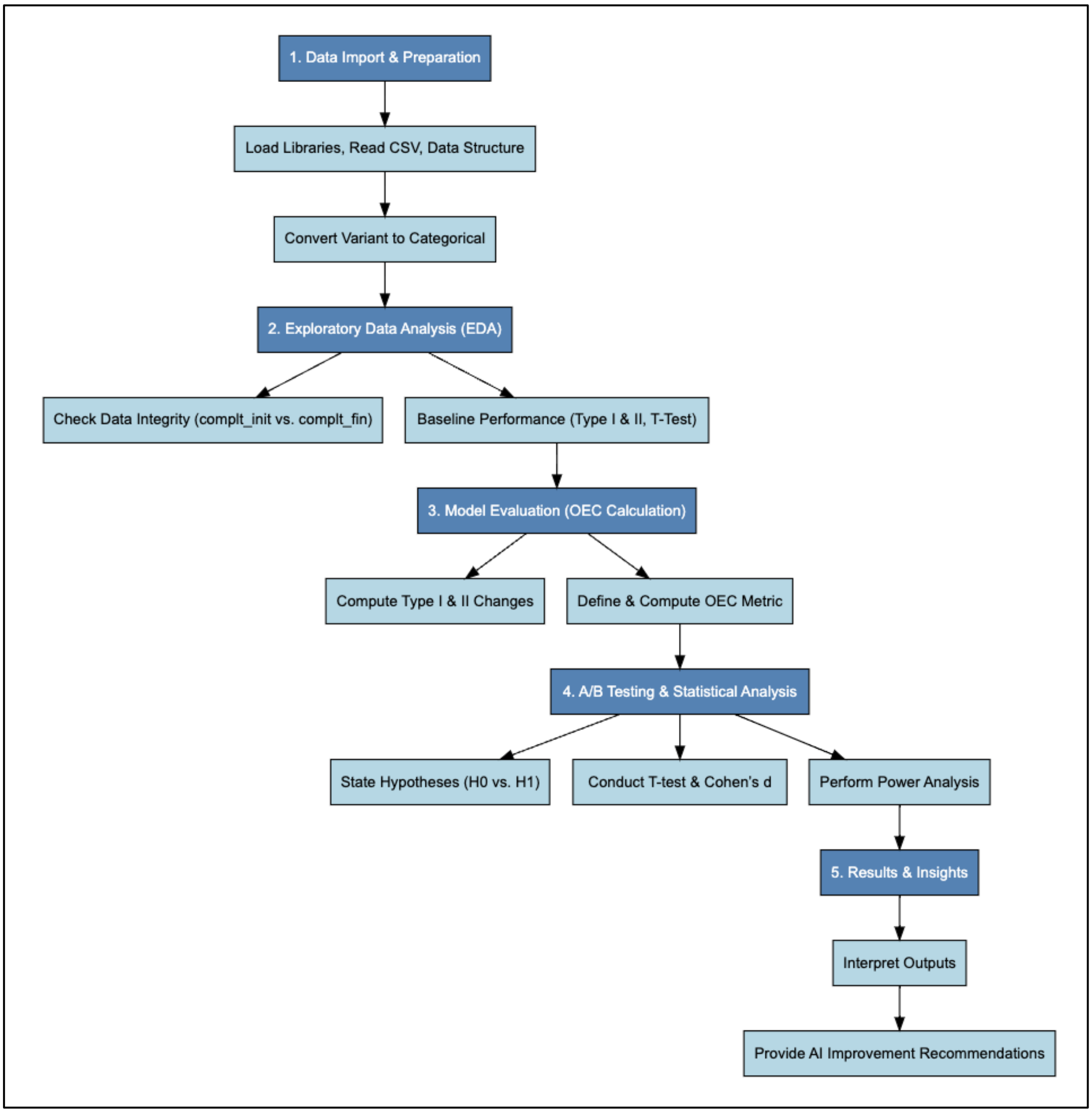
Every business has the goal to maximise the profits and minimize the losses to increase the chances of long-term success (Doyle, 2023). This project has been made to evaluate if the new model is more effective in improving the decision-making ability of the loan officers or not.

METHODOLOGY

The procedure of Preparing and Analysing the Data

This project uses A/B testing to analyse whether the newly developed AI loan approval model can improve the decision of loan officers. A/B testing is a widely adapted practice that empowers data-driven decision-making through controlled experiments by the direct comparison of two different versions of a system in the form of a test key (Quin, et al., 2024). Steps to prepare and analyse the data are given in Figure 1.

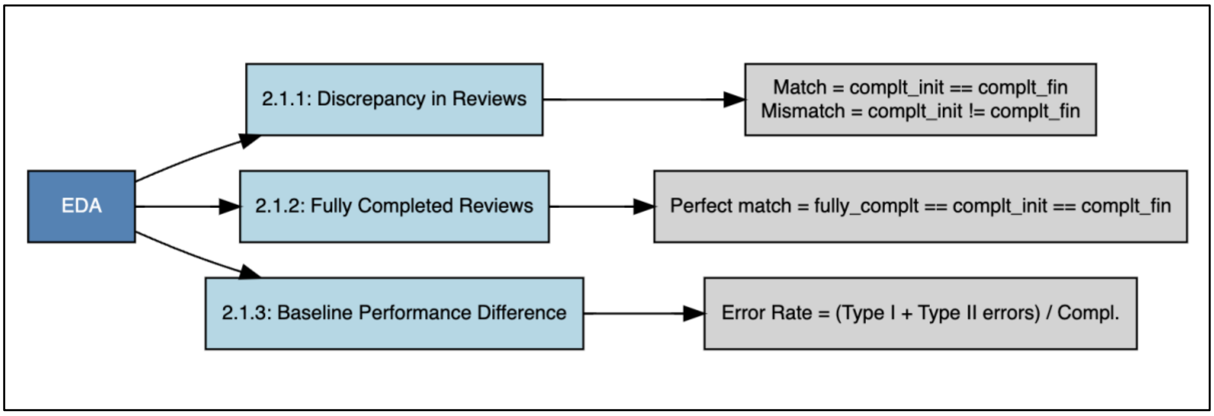
Figure 1 Process for A/B Testing



Source 1 Author's Calculations

Step 1: Exploratory Data Analysis

Figure 2 Flow Chart of EDA



Source 2 Author's Calculations

Raw data generally has missing, inconsistent and a lot of redundant information (Miller, 2019). The current data based on formula in Figure 2 set shows that the two-stage loan review process is not strictly followed.

Discrepancy in the number of loans reviewed:

In 138 out of 470 cases, the number of applications reviewed before and after AI assistance differs. Eventually, when the initial and final number of reviewed applications match, there is no confirmation whether they are the same loans, introducing uncertainty into our comparison. This means we have 96 days in total when more applications were checked in the initial stage than when seeing the model, and there are 42 days when more decisions have been made using the AI model than initial checks conducted.

Fully complete data points:

It was found that there are 332 data points where initial, final and completed reviews were congruent *(complt_init = complt_fin = fully_complt),* i.e. only 70% of the instances had fully complete data for the respective keys, leaving a significant chunk of 30% up to ambiguity.

Baseline differences in officer performance:

The small data set with 470 data points, imply that the results can be easily skewed with officer or algorithmic biases in both stages (Akter, et al., May 2022). Therefore, our aim is to achieve a **randomised controlled experiment**.

If the treatment group does not start with better decision-making quality than the control group in stage 1, then any improvement in their performance after seeing the model might not be solely due to the AI model—it could be that they were better decision-makers to begin with. To focus on overall performance differences rather than daily fluctuations, we aggregate data at the loan officer level.

Inference: Baseline performance

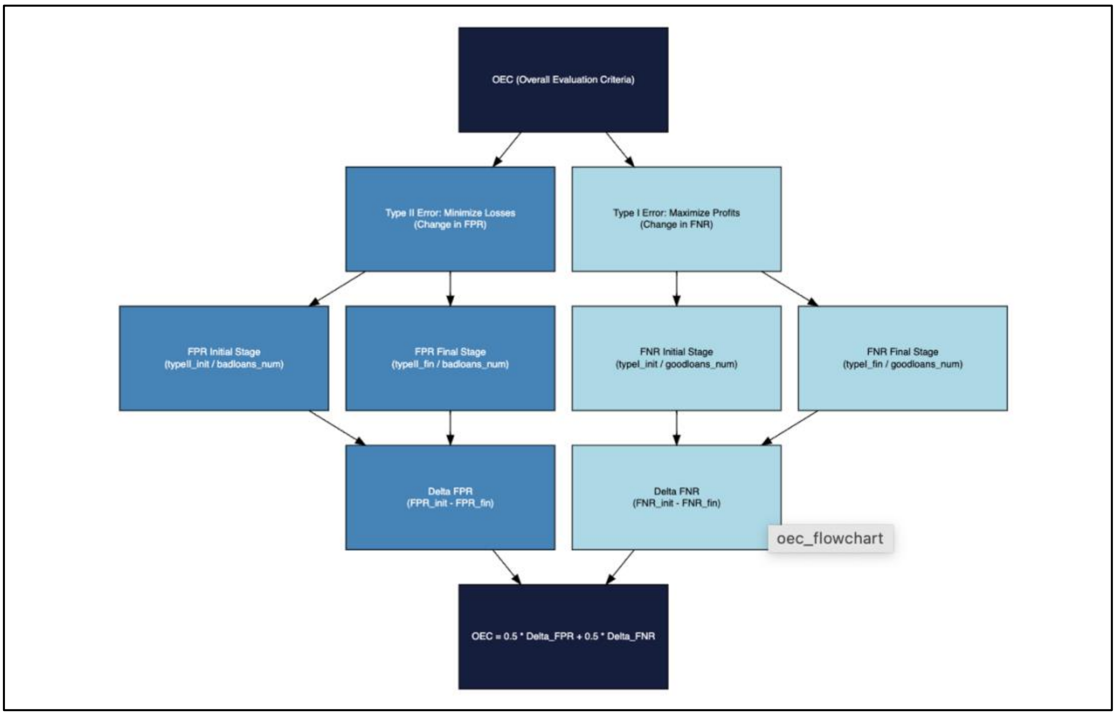
A two-sample t-test was conducted to compare the initial error rate between the Control and Treatment groups. Since the *p*-value is less than 0.05 at a 95% of confidence interval, we conclude that there is a statistically significant difference in stage 1 or initial error rate in treatment and control groups with $t(42.06) = 4.11$, for the mean difference of [0.351, 0.463]. With Cohen's $d = 1.19$ indicating a large effect size ($d > 0.8$), this empowers the practical significance of this difference.

Therefore, this pre-existing disparity complicates causal inference. Improvements in the Treatment group post-AI may not solely be due to AI assistance but rather their inherently superior decision-making skills.

Step 2: Overall Evaluation Criteria

Based of company's objective of maximising the profits and minimising the losses, the OEC for this project has 2 equally weighted parts as shown in Figure 3.

Figure 3 OEC Calculation Process



Source 3 Author's Calculations

Minimum Losses:

This metric focuses on Type II errors i.e. minimising the false positives in the data set. This means a greater change between the initial and final stages in proportion to bad loans implies a reliable outcome for the model.

Maximize Losses:

This metric focuses on Type I errors i.e. minimising false negatives in the dataset. This is calculated on the basis of minimising the rejections of good loans. Similar to the minimum loss approach, we evaluate errors in proportion to good loans. This shows that if the decision quality of the officer has gone up or down in stages 1 and 2.

Step 3: A/B Testing

Hypothesis

<u>H_0 (Null Hypothesis):</u> There is no significant difference between loan officers' performance in treatment and control groups.

<u>H_1 (Alternative Hypothesis):</u> A significant difference between loan officers' performance in treatment and control.

Since the company aims to assess overall officer performance, the data is aggregated at the **loan officer level** to compare average performance across the Control and Treatment groups.

T-Test

Table 1 Results of T-Test

Statistic	Value
t-value	2.4826
Degrees of Freedom (df)	20.226
p-value	0.02193
95% Confidence Interval	[0.0232, 0.2658]
Mean (Control Group)	0.2099
Mean (Treatment Group)	0.0654

Source 4 Author's calculation based on R code

The Welch Two Sample T-Test (Table 1) shows that there is a statistically significant between the control and treatment groups, $t(20.226)=2.483$, $p=0.022$, 95% CI[0.0232, 0.266], which implies that the true mean difference lies in the range of 0.023 and 0.266. Since the mean OEC score in the control group (M = 0.21) is 0.14 units higher than in the treatment group (M = 0.07), this suggests that the new AI model may not be improving decision-making quality as expected.

Therefore, we reject the null hypothesis and conclude that loan officers using the AI model performed worse in minimising errors (Type I and Type II), suggesting a decline in overall decision-making quality.

Step 4: Checking the Effect Size

Table 2 Results of Effect Size- Cohen's D Method

Statistic	Value
Cohen's d	-0.87
95% CI Lower Bound	-1.48
95% CI Upper Bound	-0.26

Source 5 Author's calculation based on R code

Due to low data integrity, as caveated in Step 1: **Exploratory Data Analysis**, the control group is marginally better than the treatment, merely passing the benchmark of $p<0.05$.

The calculated Cohen's d = -0.87, with a 95% confidence interval [-1.48, -0.26], indicates a large negative effect size (Table 2). With a negative effect size, it can be said for certain that the treatment group performed worse than the control group, further backing the concerns about the new model's effectiveness in improving the baseline performances of the loan officers.

Step 5: Running the Power Analysis

Table 3 Results of Power Analysis

Parameter	Value
Sample Size (n)	63.76
Cohen's d	0.5
Significance Level (α)	0.05
Power	0.8
Alternative Hypothesis	Two-sided

Source 6 Author's calculation based on R code

In this power analysis (Table 3), we choose Cohen's d = 0.5 as the effect size based on Cohen (1988). A medium effect size represents a moderate but meaningful difference, making it a reasonable choice for business and behavioural experiments. In our case, we assume that the new AI model will noticeably improve decision-making quality, but not drastically, which justifies our selection of this effect size.

The power analysis result shows that to detect a medium effect size (d = 0.5) with 80% power at a 5% significance level.

RECOMMENDATIONS

Experiment Continuation

Recommendation: The new experiment should *not continue* running and be *stopped*.

When tests continue to run after a significant outcome is not in the favour of a researcher's bias, the integrity of the A/B testing framework is compromised (Sando, 2020). In this case, based on the t-test results where the old variant is better than the new computer model OEC $t(20.226)=2.483$, $p=0.022$, 95% CI[0.0232, 0.266], it shows that even if more data points are added, there will be an imbalance of variant class and make the analysis inaccurate (Olamendy, 2024). It is evident from Cohen's recommendation of small, medium and large *d* values that the new computer model has a *large* negative effect *(-0.87)* on decision-making quality.

Evaluation of Current Experiment Design

Recommendation: The new experiment doesn't have a better performance and lacks external validity, a *revised design is necessary*.

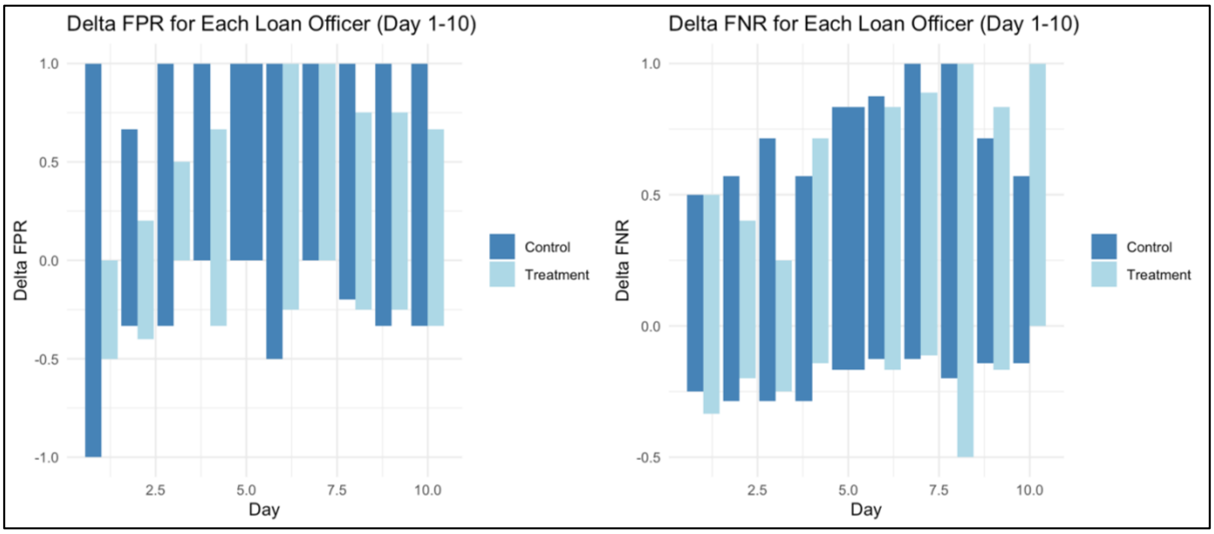
The current design has led to a statistically significant but practically worse performance in the treatment group. According to Cohen's power analysis, there should be 63 officers in each group. This means that with a total sample size of 126 loan officers (Control + Treatment), we will have 80% probability of correctly rejecting the null hypothesis and coming to a practically strong conclusion.

CONCLUSION

AI models can enhance SME's decision-making, but when experts are not referring to it proficiently, the potential for additional improvement is limited.

The Figure 4 shows that the treatment group's FNR exhibits higher variability between initial and final decisions. A limitation of the study is that on Day 5, the sum of Delta FPR and Delta FNR for the treatment group is 0, suggesting an exact balance between positive and negative changes. Overall, the higher variance in the treatment group suggests less consistency in decision-making compared to the control group at the day-level analysis.

Figure 4 Delta Over 10-Days



Source 7 Author's Calculations

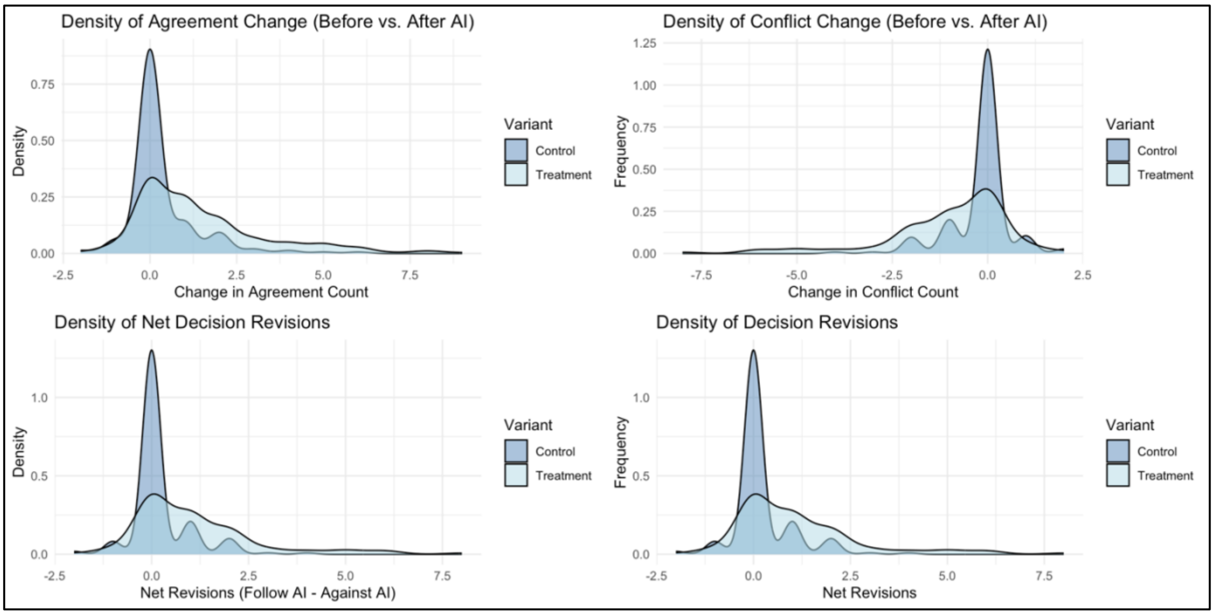
Limitations of the OEC

The OEC calculated by, bad loans approvals(Type II errors) leads to financial losses whereas rejecting good loans (Type I errors) results in missed profit but no direct loss. The model revision metric is possible with a feature importance weight via Random Forest or calculating F1 score. However to simplify, current OEC uses equal weightage for both error types, implying that avoiding both rejecting good loans and approving bad loans is equally important.

Modifications for a Better Experiment Design

A major area for improvement is Conflict resolution. The Figure 5 indicates that, while the Treatment group had a wider variety of changes in conflict, AI sometimes seemed to be the cause of these disagreements.

Figure 5 Agreement-Conflict Density Plots



Source 8 Author's Calculations

The balance of AI influence on decision revisions in the Treatment group seemed to show a much broader spread regarding decision revisions; hence, model was either being somewhat too persuasive or was underused. There should be additional training or guidelines to loan officials that would inform them when to override the AI to prevent complete dependence while assuring them that AI is nothing but a tool to aid human judgment.

Feature importance by Random Forest, GBDT, etc to enhance the experience and preferences of loan officers can help in inducing a relevant and balanced approach. Given that the Treatment group has been observed to be more variable in terms of responses with AI, a customizable AI interface could be provided whereby an officer can decide how much influence AI could have on his or her decision-making with respect to personal experience or loan complexity.

A monitoring and feedback loop with a model that learns from itself, like a neural network, will be ideal for perfecting AI toward the ever-changing needs of officers (Shah & Jha, 2023). Given that the difference between the Control and Treatment groups was indeed statistically significant, it would be very important to keep checking on AI's effect on decision making.

Overall, the project should stop running in its current stage. The company will benefit more if the AI model, with a more intuitive, advanced, and nuanced holistic approach, is implemented with accurate support and resources to sustain large computational models.

REFERENCES

Campbell, D., Loumioti, M. & Wittenberg-Moerman, R., 2019. Making sense of soft information: interpretation bias and loan quality. *Journal of Accounting and Economics,* , Volume 68, (Issues 2–3), p. doi.org/10.1016/j.jacceco.2019.101240.

Oyeniyi, L. D. et al., 2022. Addressing behavioral biases in financial decision-making: Effective risk management tools for banking institutions. *Magna Scientia Advanced Research and Reviews,* 5(1), pp. 069-075 https://doi.org/10.30574/msarr.2022.5.1.0039.

Quin, F., Weyns, D., Galster, M. & Silva, C. C., 2024. A/B testing: A systematic literature review. *Journal of Systems and Software,* Volume 211, p. https://doi.org/10.1016/j.jss.2024.112011.

Amato, A., Osterrieder, J. R. & Machado, M. R., 2024. How can artificial intelligence help customer intelligence for credit portfolio management? A systematic literature review. *International Journal of Information Management Data Insights,* 4(2), p. https://doi.org/10.1016/j.jjimei.2024.100234.

Miller, R., 2019. *Data Preprocessing: what is it and why is important.* [Online] Available at: https://ceoworld.biz/2019/12/13/data-preprocessing-what-is-it-and-why-is-important/
[Accessed 11 February 2025].

Akter, S. et al., May 2022. Algorithmic bias in machine learning-based marketing models. *Journal of Business Research,* Volume Volume 144, pp. Pages 201-216 https://doi.org/10.1016/j.jbusres.2022.01.083.

Doyle, M., 2023. *Profit Maximization: Definition and Strategies for Business Success.* [Online]
Available at: https://www.americanexpress.com/en-us/business/trends-and-insights/articles/profit-maximization-definition-and-strategies-for-business-success/
[Accessed 11 February 2025].

Sando, D., 2020. *Extending A/B tests to reach significance: the trap even the smartest people fall into.* [Online]

Available at: https://dominicsando.medium.com/extending-a-b-tests-to-reach-significance-the-trap-even-the-smartest-people-fall-into-904f86a18f4b
[Accessed 11 February 2024].

Olamendy, J. C., 2024. *Practical ML: Addressing Class Imbalance.* [Online] Available at: https://medium.com/@juanc.olamendy/practical-ml-addressing-class-imbalance-25c4f1b97ee3
[Accessed 11 February 2025].

APPENDIX

Appendix A: List of Figures

Appendix B: List of Tables

Appendix C: R Code

````
---
title: "ADA_Assignment"
author: "Group 26"
date: "2025-02-11"
output: html_document
editor_options:
  chunk_output_type: console
---

```{r setup, include=FALSE}
knitr::opts_chunk$set(echo = TRUE)

```



``` {r}

# libraries used
library(tidyverse)
library(gridExtra)
library(lubridate)
library(kableExtra)
library(Hmisc)
library(car)
library(stringr)
library(emmeans)
library(pwr)
library(effectsize)
library(ggplot2)
library(dplyr)
library(tidyr)
library(DiagrammeR)

```

``` {r}

# read in XLS

data <- read_csv("ADA_assignment.csv")

# check data structure

str(data)

# Set 'Variant' to categorical

data$Variant <- factor(data$Variant)
````
````

~~~~
```

# **2. Exploratory Data Analysis (EDA)**
## **2.1 Data Integrity

### **1.1.1 Discrepancy between initial and final review stage**

``` {r}

# 'cmplt_init' and 'complt_fin' do not always equal
# 'cpmplt_init' vs 'cmplt_fin' & relocate 'fully_complt'

data <- data %>%
  mutate(complt_init.vs.complt_fin = ifelse(complt_init == complt_fin, "Match",
"Mismatch")) %>%
  relocate(fully_complt, .after = complt_fin)

# display matches and mismatches

table(data$complt_init.vs.complt_fin)

# differences between inital loan and final loan checks

data <- data %>%
  mutate(complt_match = case_when(
    complt_init < complt_fin ~ "Initial smaller",
    complt_init > complt_fin ~ "Initial larger",
    complt_init == complt_fin ~ "Match"
  )) %>%
  relocate(complt_match, .after = complt_fin)

table(data$complt_match)

```

---

### **1.1.2 Truly fully completed review process ('fully_complt' = 'complt_init' =
'complt_fin')**

``` {r}

data <- data %>%
  mutate(perfect_match = ifelse(fully_complt == complt_init & fully_complt ==
complt_fin, "Match", "Mismatch")) %>%
  relocate(perfect_match, .after = fully_complt)
~~~~

display matches and mismatches

table(data$perfect_match)

```

---

### **1.1.3 Baseline differences in officer performance (Control vs. Treatment)**

``` {r}

# create new subset for comparison of 'initial error rate' between control and treatment

init_data <- data %>%
  group_by(Variant, loanofficer_id) %>%
  summarise(
    typeI_init = sum(typeI_init, na.rm = TRUE),
    typeII_init = sum(typeII_init, na.rm = TRUE),
    complt_init = sum(complt_init, na.rm = TRUE),
    .groups = "drop"
  ) %>%
  mutate(err_rate_init_avg = (typeI_init + typeII_init) / complt_init)

# conduct t-test

t_test_1 <- t.test(err_rate_init_avg ~ Variant, data = init_data, var.equal = FALSE)
t_test_1

# calculate effect size using Cohen's d

cohen_d_1 <- cohens_d(err_rate_init_avg ~ Variant, data = init_data)
cohen_d_1

# means control vs treatment

(mean_control_init <- mean(init_data$err_rate_init_avg[init_data$Variant == "Control"]))
(mean_treatment_init <- mean(init_data$err_rate_init_avg[init_data$Variant == "Treatment"]))

### **1.1.4 Days without bad loans**

``` {r}

# calculate 'no bad loans days'

no_bad_loans_days <- data %>% filter(badloans_num == 0)
count(no_bad_loans_days)

data <- data %>% filter(badloans_num > 0)

```

**1.1.5 Days without AI use or officer input

``` {r}

data.frame(
  Condition = c("Only Officer Decisions", "Only Model Decisions"),
  Count = c(
    sum(data$complt_init != 0 & data$complt_fin == 0),
    sum(data$complt_init == 0 & data$complt_fin != 0)
  )
)

```

**3.2 Decision

---

3. Creating an OEC (Overall Evaluation Criteria)

1. The model's ability to help minimize losses:

$(typeII_init / badloans_num) - (typeII_fin / badloans_num)$ (Delta_FPR)

2. The model's ability to help maximize profits:
$(typeI_init / goodloans_num) - (typeI_fin / badloans_num)$ (Delta_FNR)

This results in the final expression as follows:

$OEC = 0.5 * Delta_FPR + 0.5* Delta_FNR$

``` {r}
```

```
# Type II error to minimize costs

data <- data %>%
  mutate(FPR_init = typeII_init / badloans_num) %>%
  mutate(FPR_fin = typeII_fin / badloans_num) %>%
  mutate(Delta_FPR = FPR_init - FPR_fin)

# switch column position

data <- data %>%
  relocate(FPR_init, .after = goodloans_num) %>%
  relocate(FPR_fin, .after = FPR_init) %>%
  relocate(Delta_FPR, .after = FPR_fin)

# Type I error to maximize profits

data <- data %>%
  mutate(FNR_init = typeI_init / goodloans_num) %>%
  mutate(FNR_fin = typeI_fin / goodloans_num) %>%
  mutate(Delta_FNR = FNR_init - FNR_fin) %>%
  relocate(Delta_FNR, .after = Delta_FPR)
```

``` {r}

# Final OEC

data <- data %>%
  mutate(OEC = 0.5 * Delta_FPR + 0.5 * Delta_FNR) %>%
  relocate(OEC, .after = Delta_FNR)

```

4. A/B Testing
4.1 Hypothesis

H0: There is no difference in loan officers' performance between those using the old AI model and those using the new model.

H1: There is a difference in loan officers' performance between those using the old AI model and those using the new model.

``` {r}

# aggregating data at loan officer level
```

loan_officer_level <- data %>%
  group_by(Variant, loanofficer_id) %>%
  summarise(across(where(is.numeric), mean), .groups = "drop")

```

## **4.2 T-test**

``` {r}

# run two-sample t-test estimating the difference in mean between control and treatment

OEC_t.test_result <- t.test(OEC ~ Variant, data = loan_officer_level, var.equal = FALSE)
OEC_t.test_result

```

4.3 Cohen d

``` {r}

# using Cohen's measure to compute & interpret effect size
# absolute d value of 0.2 or less: small, 0.3-0.5 medium, 0.8 or bigger: large effect size

Control = loan_officer_level$OEC[loan_officer_level$Variant == "Control"]
Treatment = loan_officer_level$OEC[loan_officer_level$Variant == "Treatment"]
cohens_d(Treatment, Control) # compute effect size of difference between Treatment 1 & Control

```

4.4 Power Analysis

``` {r}

# power analysis to estimate required sample size (number of loan officers) for new, redesigned experiment

pwr.t.test(power = .8, # 80% power
d = .6, # Cohen's d
sig.level = 0.05, # threshold for p-val
type = "two.sample")

```

5 Plots
```

```{r}
# Data Preparation: Calculate net agreement/conflict changes and net revisions
data <- data %>%
  mutate(
    agree_change = agree_fin - agree_init,
    conflict_change = conflict_fin - conflict_init,
    revision_net = revised_per_ai - revised_agst_ai
  )

# 1. Density Plot for Agreement Changes (Treatment vs Control)
p1 <- ggplot(data, aes(x = agree_change, fill = Variant)) +
  geom_density(alpha = 0.5) +
  labs(title = "Density of Agreement Change (Before vs. After AI)",
       x = "Change in Agreement Count",
       y = "Density") +
  theme_minimal()+
  scale_fill_manual(values = c("#4683B7", "lightblue"))

# 2. Histogram for Conflict Changes (Treatment vs Control)
p2 <- ggplot(data, aes(x = conflict_change, fill = Variant)) +
  geom_density(alpha = 0.5)+
  labs(title = "Density of Conflict Change (Before vs. After AI)",
       x = "Change in Conflict Count",
       y = "Frequency") +
  theme_minimal()+
  scale_fill_manual(values = c("#4683B7", "lightblue"))

# 3. Density Plot for Decision Revisions (Treatment vs Control)
p3 <- ggplot(data, aes(x = revision_net, fill = Variant)) +
  geom_density(alpha = 0.5) +
  labs(title = "Density of Net Decision Revisions",
       x = "Net Revisions (Follow AI - Against AI)",
       y = "Density") +
  theme_minimal()+
  scale_fill_manual(values = c("#4683B7", "lightblue"))

# 4. Histogram for Decision Revisions (Treatment vs Control)
p4 <- ggplot(data, aes(x = revision_net, fill = Variant)) +
  geom_density(alpha = 0.5) +
  labs(title = "Density of Decision Revisions",
       x = "Net Revisions",
       y = "Frequency") +
  theme_minimal()+
  scale_fill_manual(values = c("#4683B7", "lightblue"))

# Arrange all graphs in one layout
grid.arrange(p1, p2, p3, p4, ncol = 2)


```

````
```{r}

# Ensure all days & Variants are included, even if no data exists
data <- data %>%
  complete(day, Variant, fill = list(Delta_FPR = 0, Delta_FNR = 0))

# Plot for Delta FPR
p1 <- ggplot(data, aes(x = as.factor(day), y = Delta_FPR, fill = Variant)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Delta FPR for Each Loan Officer (Day 1-10)",
       x = "Day", y = "Delta FPR") +
  theme_minimal() +
  scale_fill_manual(values = c("#4683B7", "lightblue")) +
  theme(legend.title = element_blank())

# Plot for Delta FNR
p2 <- ggplot(data, aes(x = as.factor(day), y = Delta_FNR, fill = Variant)) +
  geom_bar(stat = "identity", position = "dodge") +
  labs(title = "Delta FNR for Each Loan Officer (Day 1-10)",
       x = "Day", y = "Delta FNR") +
  theme_minimal() +
  scale_fill_manual(values = c("#4683B7", "lightblue")) +
  theme(legend.title = element_blank())

# Arrange the plots
grid.arrange(p1, p2, ncol = 2)

```
```{r}

```


```{r}
# Install and load the required package
# install.packages("DiagrammeR")
library(DiagrammeR)

# Define the flowchart structure with color adjustments
flow_chart <- "
digraph oec_flowchart {
  node [shape=box, style=filled, fontname='Helvetica', fontsize=10, width=3,
height=1.5]

  # OEC heading in grey
  OEC [label='OEC (Overall Evaluation Criteria)', style=filled, fillcolor='#151e3d',
fontcolor=white, fonttype=bold]

  # Left side (Type II error) in navy blue
  typeII [label='Type II Error: Minimize Losses\n(Change in FPR)', style=filled,
fillcolor='#4683B7', fontcolor=white]
````
````

```
  FPR_init [label='FPR Initial Stage\n(typeII_init / badloans_num)', style=filled,
fillcolor='#4683B7', fontcolor=white]
  FPR_fin [label='FPR Final Stage\n(typeII_fin / badloans_num)', style=filled,
fillcolor='#4683B7', fontcolor=white]
  Delta_FPR [label='Delta FPR\n(FPR_init - FPR_fin)', style=filled,
fillcolor='#4683B7', fontcolor=white]

  # Right side (Type I error) in light blue
  typeI [label='Type I Error: Maximize Profits\n(Change in FNR)', style=filled,
fillcolor=lightblue]
  FNR_init [label='FNR Initial Stage\n(typeI_init / goodloans_num)', style=filled,
fillcolor=lightblue]
  FNR_fin [label='FNR Final Stage\n(typeI_fin / goodloans_num)', style=filled,
fillcolor=lightblue]
  Delta_FNR [label='Delta FNR\n(FNR_init - FNR_fin)', style=filled, fillcolor=lightblue]

  # Weighting and OEC calculation
  OEC_final [label='OEC = 0.5 * Delta_FPR + 0.5 * Delta_FNR', style=filled,
fillcolor='#151e3d', fontcolor=white]

  # Connect the nodes
  OEC -> typeII
  OEC -> typeI

  typeII -> FPR_init
  typeII -> FPR_fin
  FPR_init -> Delta_FPR
  FPR_fin -> Delta_FPR

  typeI -> FNR_init
  typeI -> FNR_fin
  FNR_init -> Delta_FNR
  FNR_fin -> Delta_FNR

  # Final OEC Calculation
  Delta_FPR -> OEC_final
  Delta_FNR -> OEC_final
}
"

# Create and display the flowchart
grViz(flow_chart)

```

```{r}
# Creating a horizontal flowchart for EDA
grViz("
digraph EDA_flowchart {

  # Define node attributes for styling
```

```
  node [shape=box, style=filled, fontname=helvetica, fontsize=10]

  # Main node
  EDA [label='EDA', shape=box, fillcolor='#4683B7', fontcolor=white]

  # Sub-nodes for EDA process (2.1.1 to 2.1.5)
  discrepancy [label='2.1.1: Discrepancy in Reviews', fillcolor=lightblue]
  discrepancy_formula [label='Match = complt_init == complt_fin\nMismatch =
complt_init != complt_fin']

  fully_completed [label='2.1.2: Fully Completed Reviews', fillcolor=lightblue]
  fully_completed_formula [label='Perfect match = fully_complt == complt_init ==
complt_fin']

  baseline_performance [label='2.1.3: Baseline Performance Difference',
fillcolor=lightblue]
  baseline_performance_formula [label='Error Rate = (Type I + Type II errors) /
Compl.']



  # Define edges to connect nodes
  EDA -> discrepancy
  discrepancy -> discrepancy_formula

  EDA -> fully_completed
  fully_completed -> fully_completed_formula

  EDA -> baseline_performance
  baseline_performance -> baseline_performance_formula


  # Additional styling for clarity
  edge [color=gray]
  rankdir=LR
}
")
```