RL for A/B Testing in E-commerce

Itay Toledano, Amir Yorav
204116073, 037378874

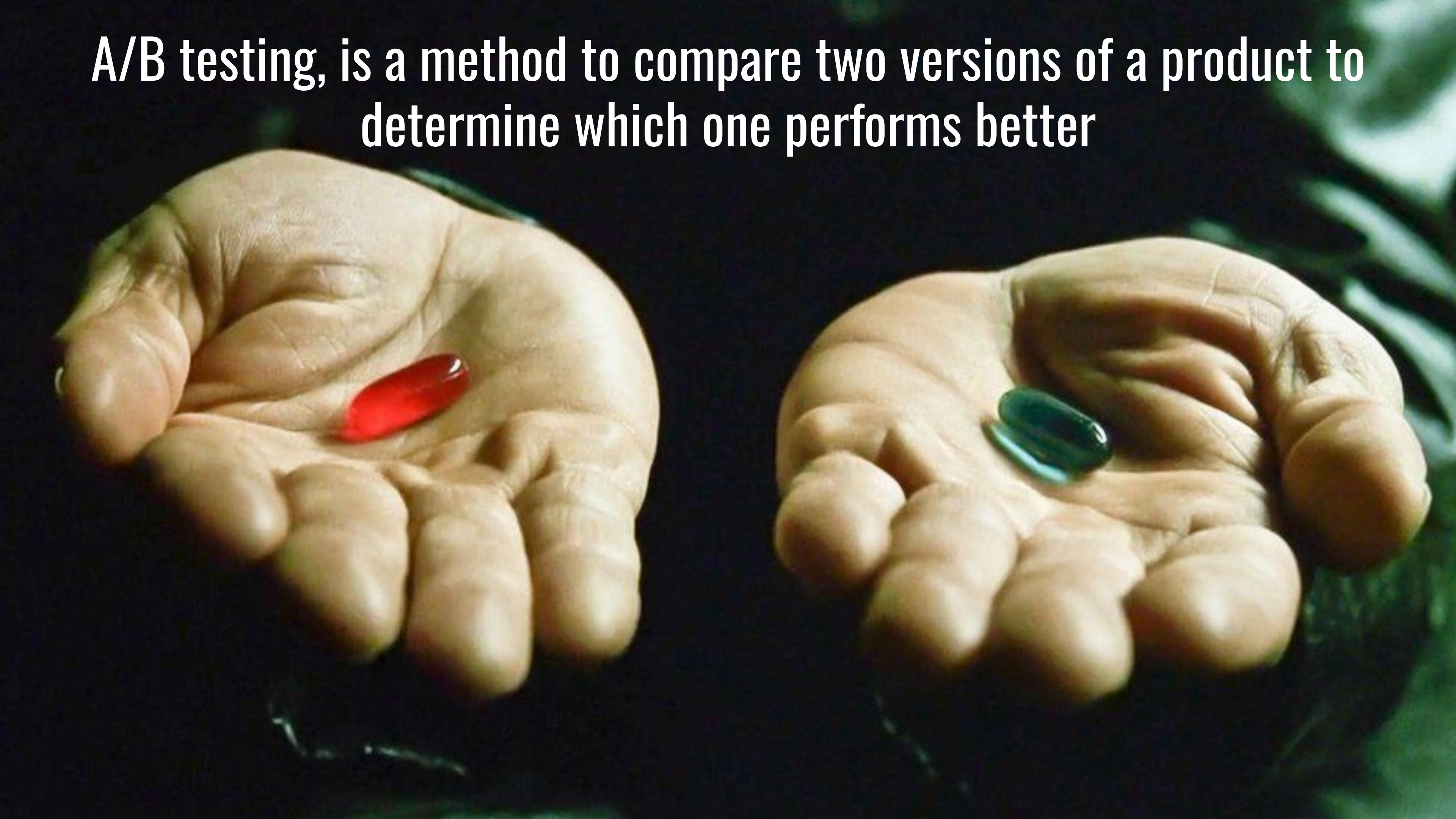
18.6.2024

Outline

- Concepts
    - A/B Testing
    - E-commerce
    - RL for A/B Testing
- Our Study
    - Dataset
    - Methods
    - Results & Analysis
- Conclusions

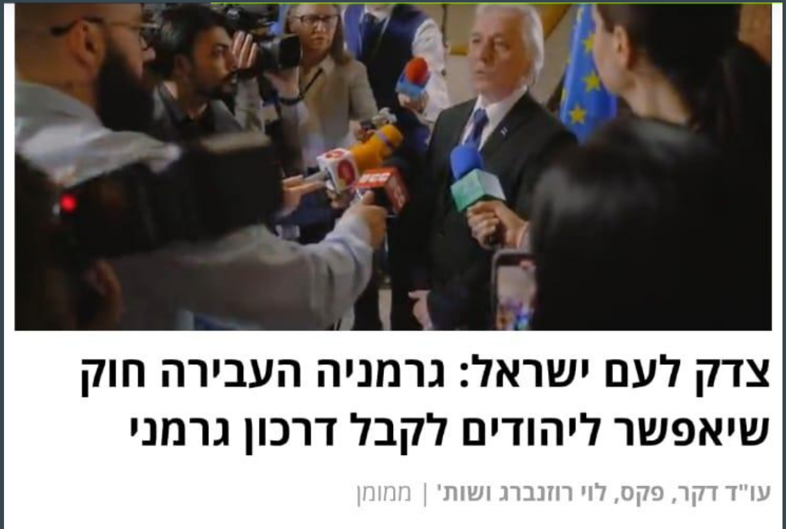
A/B Testing

A/B testing, is a method to compare two versions of a product to determine which one performs better



I'M A/B TESTING TO SEE WHAT GETS THE MOST CANDY.
TOM FISH BURNE
© marketoonist.com

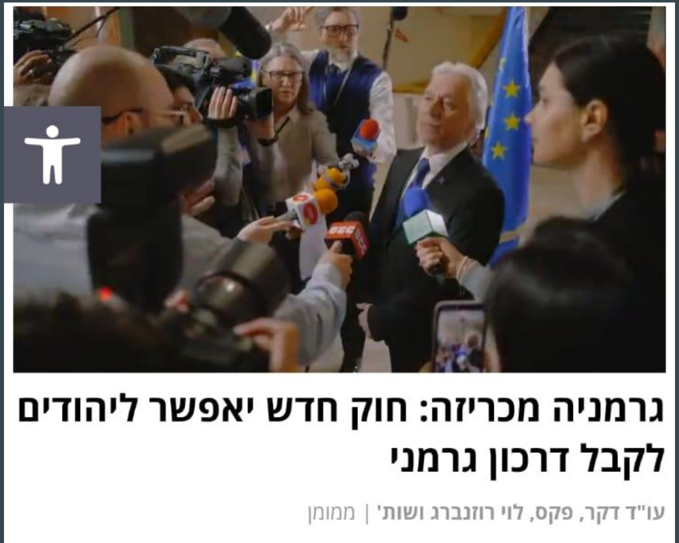
A



צדק לעם ישראל: גרמניה העבירה חוק
שיאפשר ליהודים לקבל דרכון גרמני

עו"ד דקר, פקס, לוי רוזנברג ושות' | ממומן

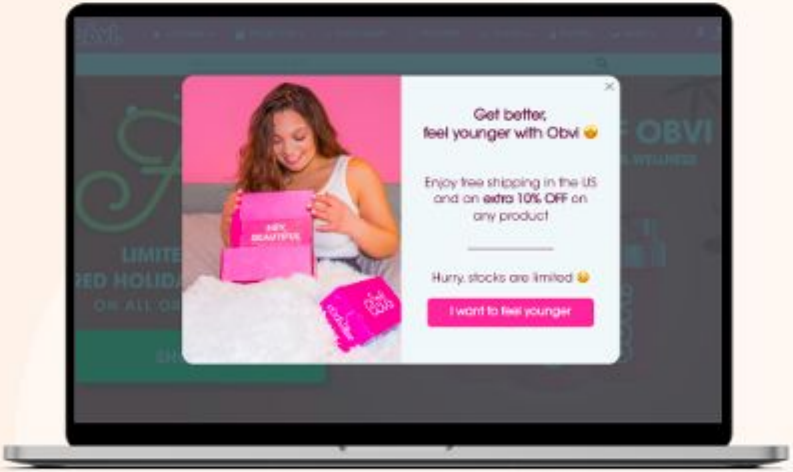
B



גרמניה מכריזה: חוק חדש יאפשר ליהודים
לקבל דרכון גרמני

עו"ד דקר, פקס, לוי רוזנברג ושות' | ממומן

+7.97% conversion rate

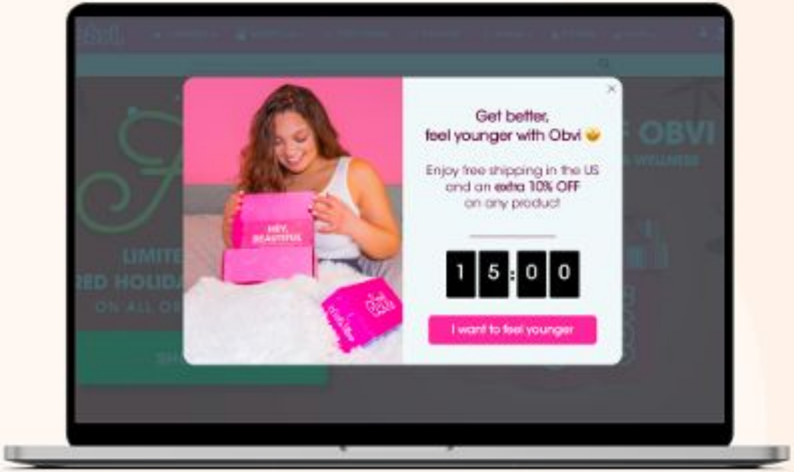


Get better, feel younger with Obvi 😊

Enjoy free shipping in the US and an **extra 10% OFF** on any product

Hurry, stocks are limited 😊

Without countdown timer

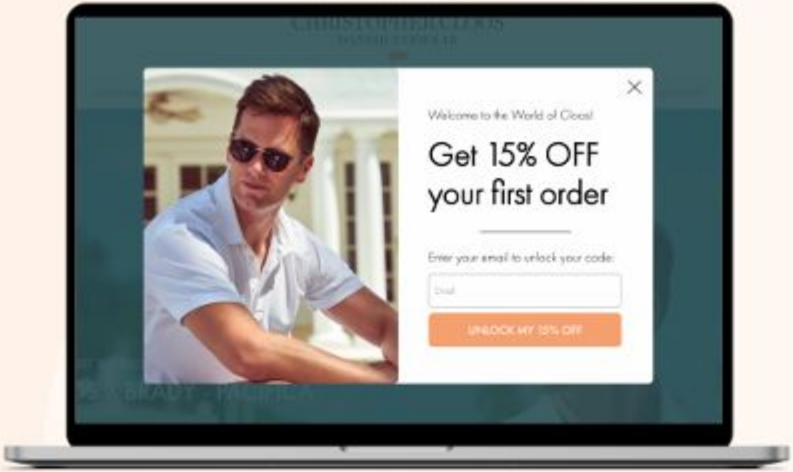


Get better, feel younger with Obvi 😊

Enjoy free shipping in the US and an **extra 10% OFF** on any product

With countdown timer

+15.38% conversion rate

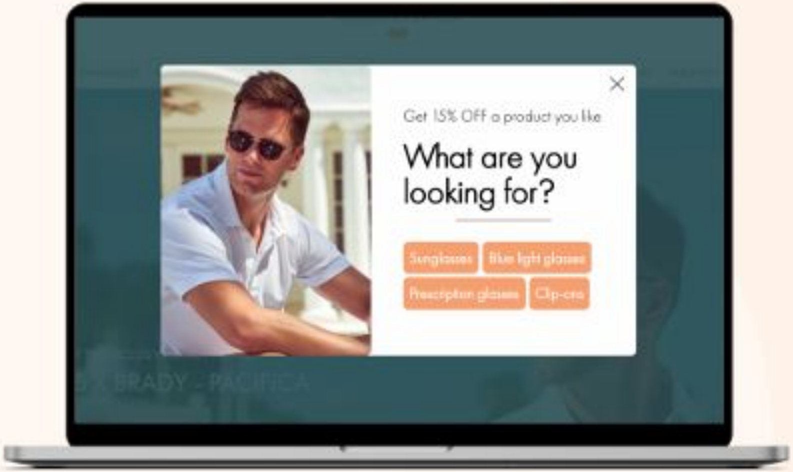


Welcome to the World of Cloos!

Get 15% OFF your first order

Enter your email to unlock your code:

Classic welcome popup



Get 15% OFF a product you like

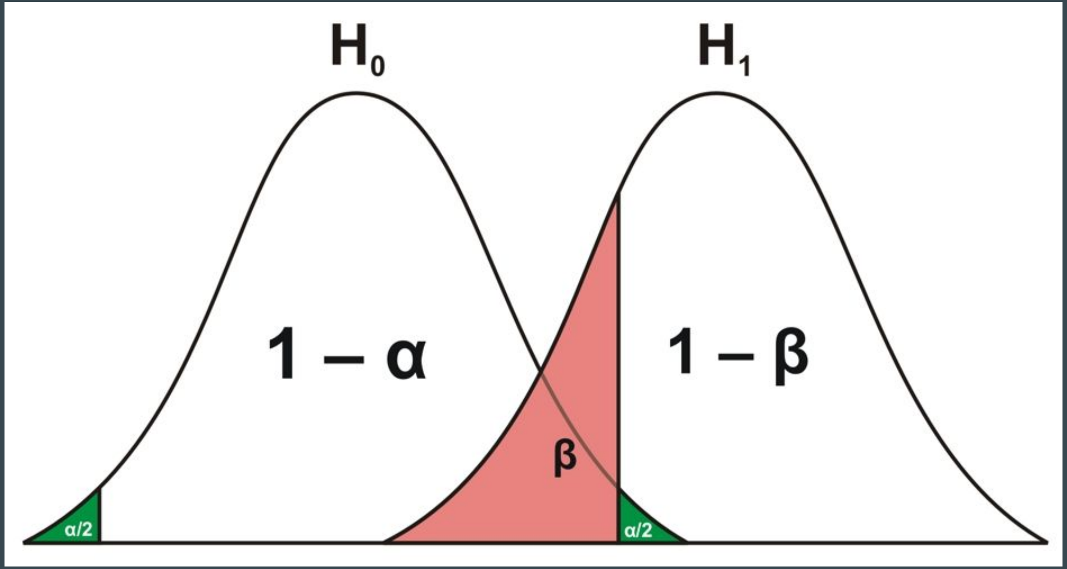
What are you looking for?

Conversational popup

MAKING DATA-DRIVEN DECISIONS

Traditional A/B Testing

1. Define the Objective
2. Formulate Hypotheses
3. Create Variations
4. Split Your Audience
5. Ensure Sample Size is Sufficient
6. Run the Experiment
7. Collect Data
8. Analyze the Results
9. Draw Conclusions
10. Make Decisions



Challenges in Traditional A/B Testing

Single-Unit Experiments: Impact of sequential treatments over time.

Dynamic Environments: Difficulty in managing changes and interactions.

Limited Scope: Often does not account for long-term effects.

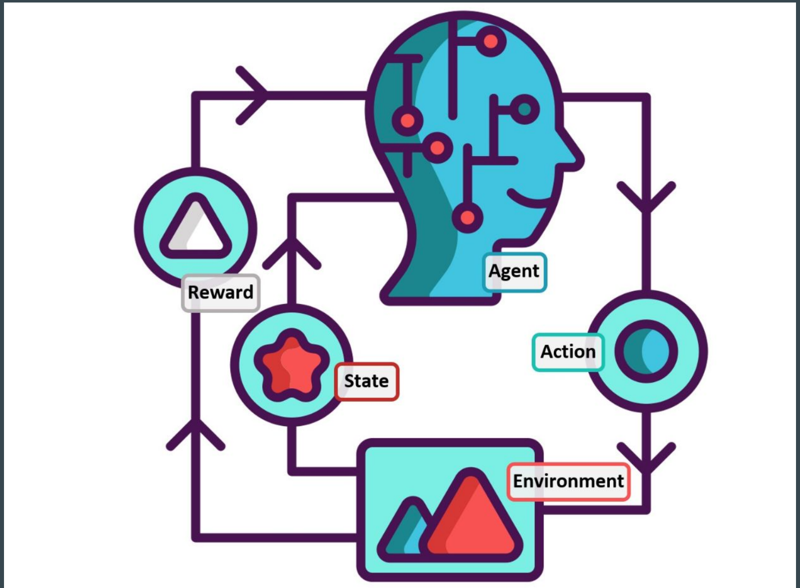
Reinforcement Learning

RL is a type of machine learning where an agent learns to make decisions by performing actions and receiving rewards.

States: Snapshot of the environment.

Actions: Choices made by the agent.

Rewards: Feedback from the environment.



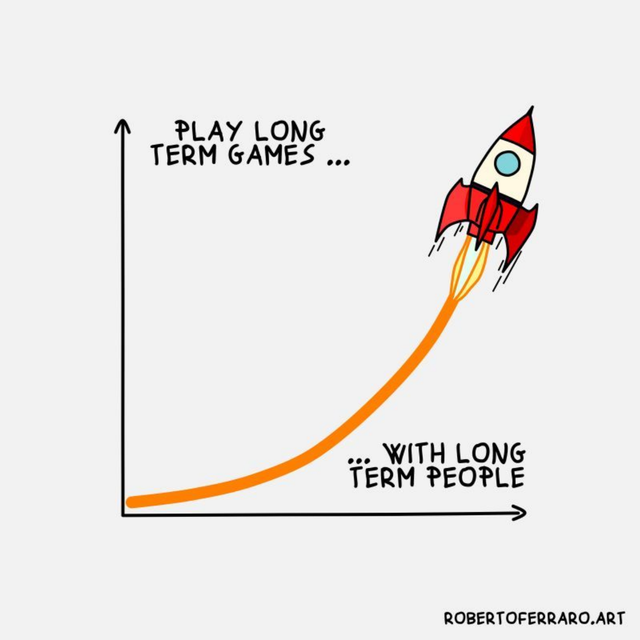
Applying RL to A/B Testing

Framework:

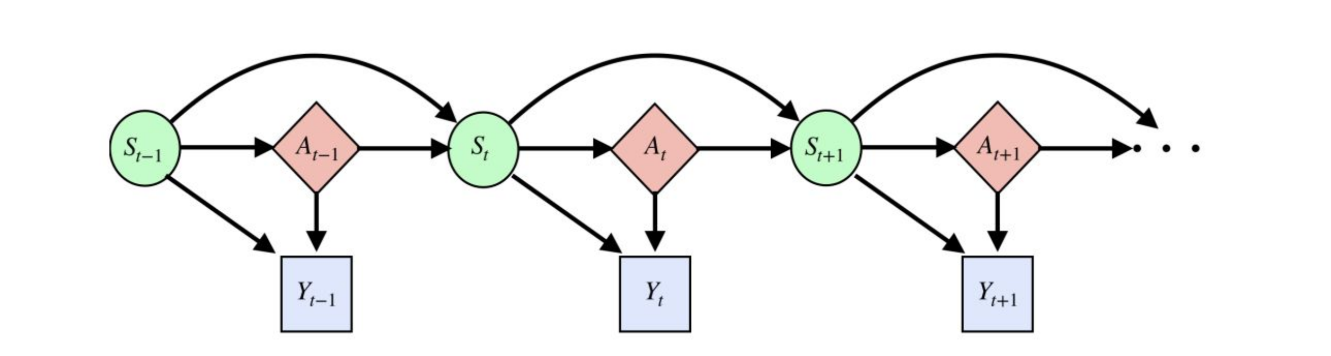
1. Model the environment and user interactions.
2. Define states, actions, and rewards.
3. Use RL algorithms to optimize long-term outcomes.

Advantages:

- ➔ Adaptive decision-making.
- ➔ Optimization for long-term effects.
- ➔ Continuous learning and improvement.



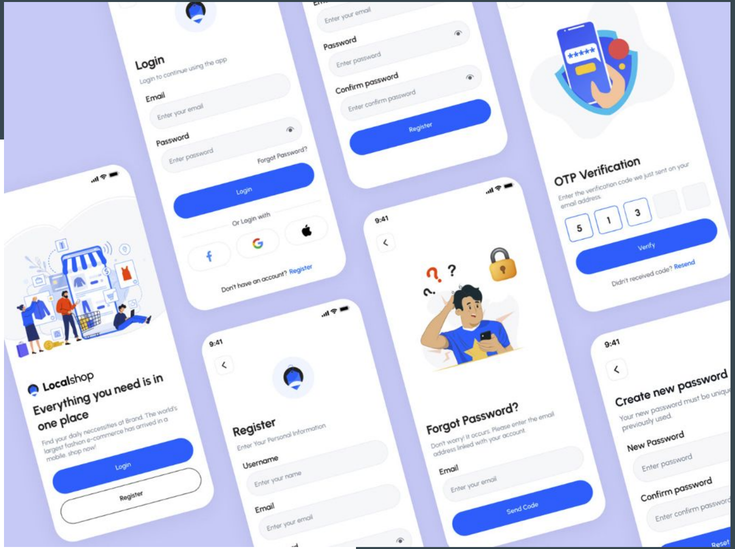
Dynamic Causal Effects Evaluation in A/B Testing with a Reinforcement Learning Framework (2023)



- Markov Decision Process
- Outcomes feed into the RL process, which updates the Value Function.
- The updated Value Function influences future actions, leading to improved decision-making

Our Study

Environment - The Onboarding page


Localshop
Everything you need is in one place
Find your daily neccesities at Brand. The world's largest fashion e-commerce has arrived in a mobile. shop now!
Login
Register
Login
Login to continue using the app
Email
Enter your email
Password
Enter password
Forgot Password?
Login
Or Login with
Don't have an account? Register
Enter your email
Password
Enter password
Confirm password
Enter confirm password
Register
9:41
Register
Enter Your Personal Information
Username
Enter your name
Email
Enter your email
9:41
Forgot Password?
Don't worry! It occurs. Please enter the email address linked with your account.
Email
Enter your email
Send Code
OTP Verification
Enter the verification code we just sent on your email address.
5
1
3
Verify
Didn't received code? Resend
9:41
Create new password
Your new password must be unique from those previously used.
New Password
Enter password
Confirm password
Enter confirm password
Reset

Referral Entry

- Users enter the site through referral URLs such as Facebook and Instagram, marking the initial contact point in the conversion funnel.

User Registration

- Users are required to fill in personal details and respond to a set of questions designed to tailor their experience and gather relevant data.

Security Verification

- Users must upload their identification documents.

Customized Deposit Offer

- Based on the information provided in the previous stages, a special offer for the first deposit is presented to the users.

Homepage Access

- Users reach the homepage, where they can fully engage with the services and content provided.

The Dataset

user_id	Unique identifier for each user
age	Age of the user
ip_location	IP-based location of the user
current_browsing_time	Time spent browsing the site during the current session
pages_visited	Number of pages visited during the current session
discount	Discount offered to the user
referral_source	Source from which the user was referred (e.g., Google, Facebook)
onboarding_time	Time taken to complete the onboarding process
kyc_type	Type of ID document uploaded (e.g., passport, driver's license)
browser_type	Browser used by the user
payment_method	Payment method chosen by the user (e.g., CC, Crypto)
time_of_day	Time of day when the user completed the onboarding
planned_monthly_deposit	User's planned monthly deposit amount
device	Device used by the user (e.g., Android, iOS)
device_model	Model of the device used
base_suggested_amount	Initial deposit amount suggested
suggested_amount_after_discount	Deposit amount after applying the discount
user_action	User's action after seeing the discount (e.g., deposit, skip)

States

Each state is a vector of user features including age, IP location, browsing time, pages visited, referral source, onboarding time, KYC type, browser type, payment method, time of day, planned monthly deposit, device, device model, and base suggested amount.

Actions

The possible actions the agent can take are offering different discounts - 5%, 10%, 15%.

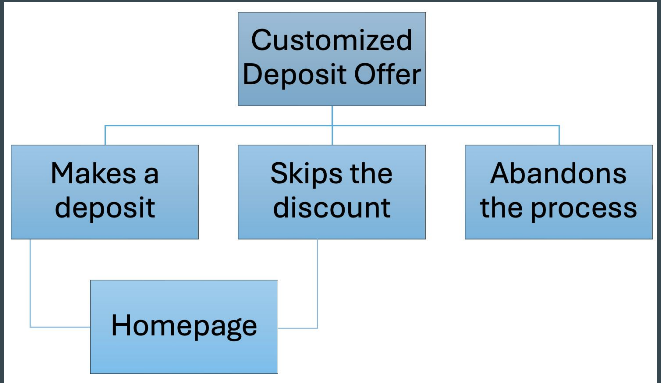
Rewards

The reward is the amount deposited by the user after the discount is applied.

a. Positive reward if the user makes a deposit.

b. Minor penalty if the user skips the discount offer.

c. Larger penalty if the user abandons the process entirely.

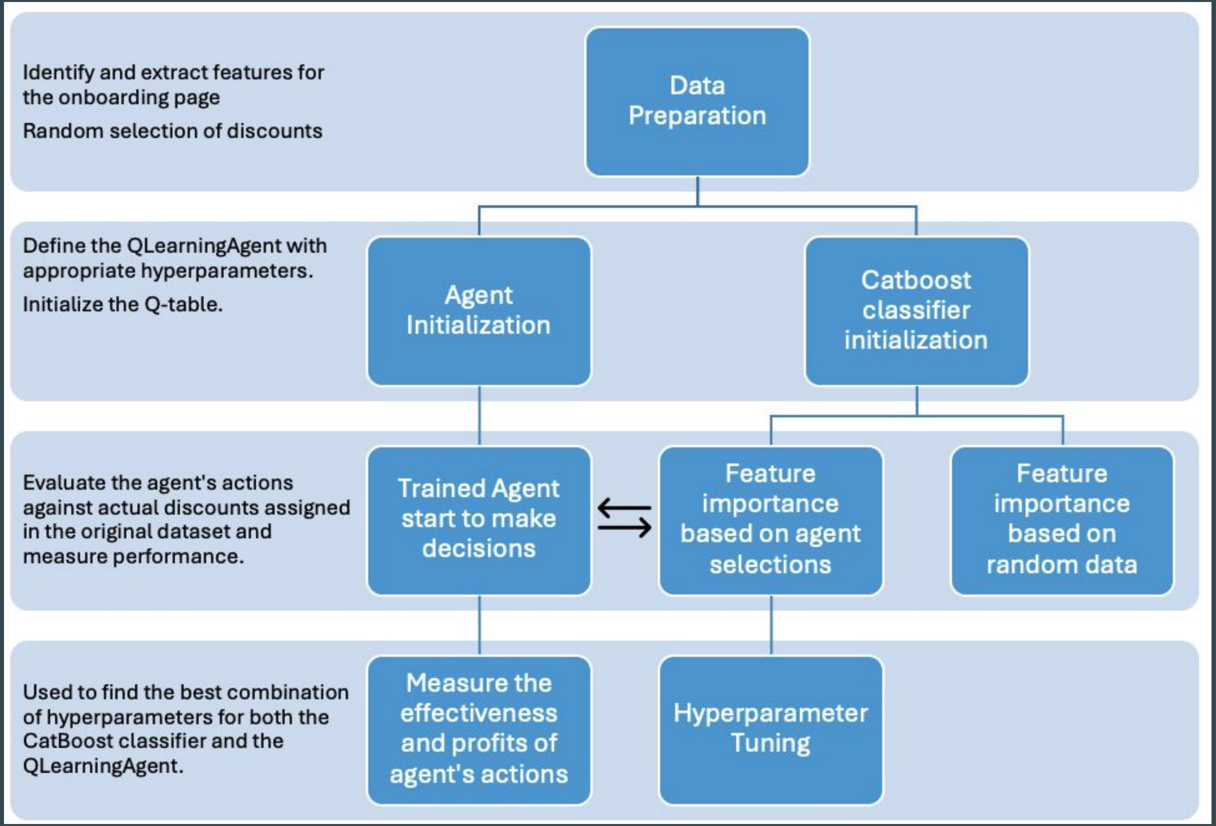


Training the Agent

The RL agent uses the Q-learning algorithm to learn the optimal discount strategy.

- Initialization: Initialize the Q-values for all state-action pairs.
- Exploration vs. Exploitation: During training, the agent explores different actions to learn their impact, balanced with exploiting known information to maximize rewards.
- Updating Q-values: Q-values are updated based on the Bellman equation, incorporating rewards received and future reward estimates.
- Regularization: L2 regularization is applied to stabilize learning and prevent overfitting.

ML-flow Architecture



Identify and extract features for the onboarding page

Random selection of discounts

Data Preparation

Define the QLearningAgent with appropriate hyperparameters.

Initialize the Q-table.

Agent Initialization

Catboost classifier initialization

Evaluate the agent's actions against actual discounts assigned in the original dataset and measure performance.

Trained Agent start to make decisions

⇆

Feature importance based on agent selections

Feature importance based on random data

Used to find the best combination of hyperparameters for both the CatBoost classifier and the QLearningAgent.

Measure the effectiveness and profits of agent's actions

Hyperparameter Tuning

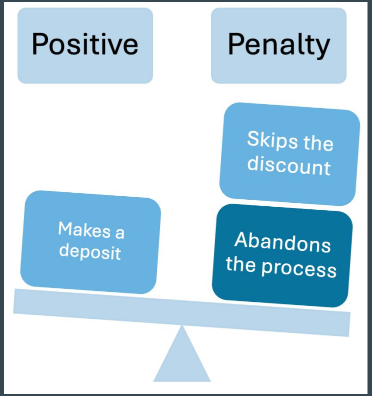
Results and Analysis

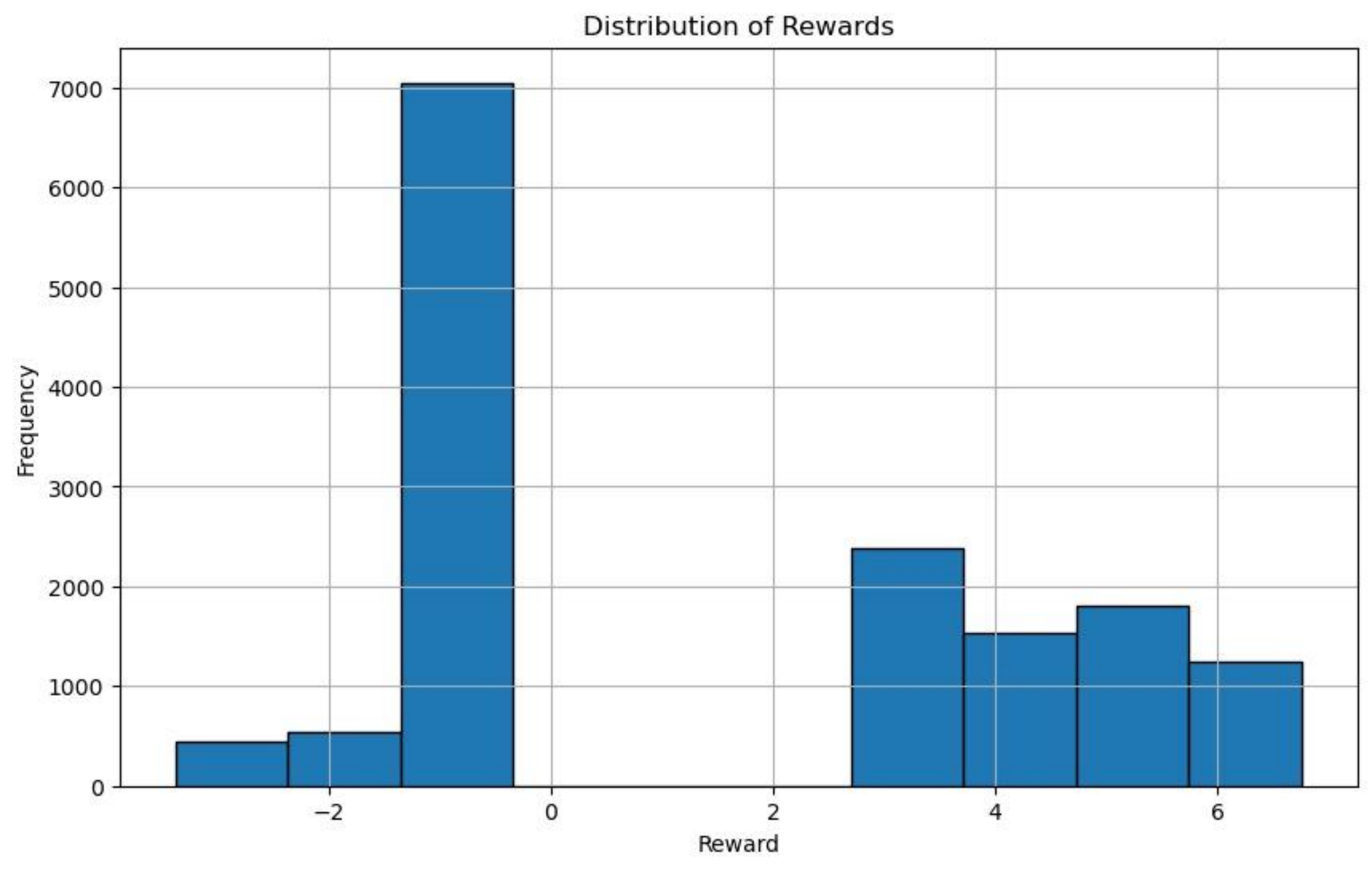
Rewards

a. Positive

b. Minor penalty

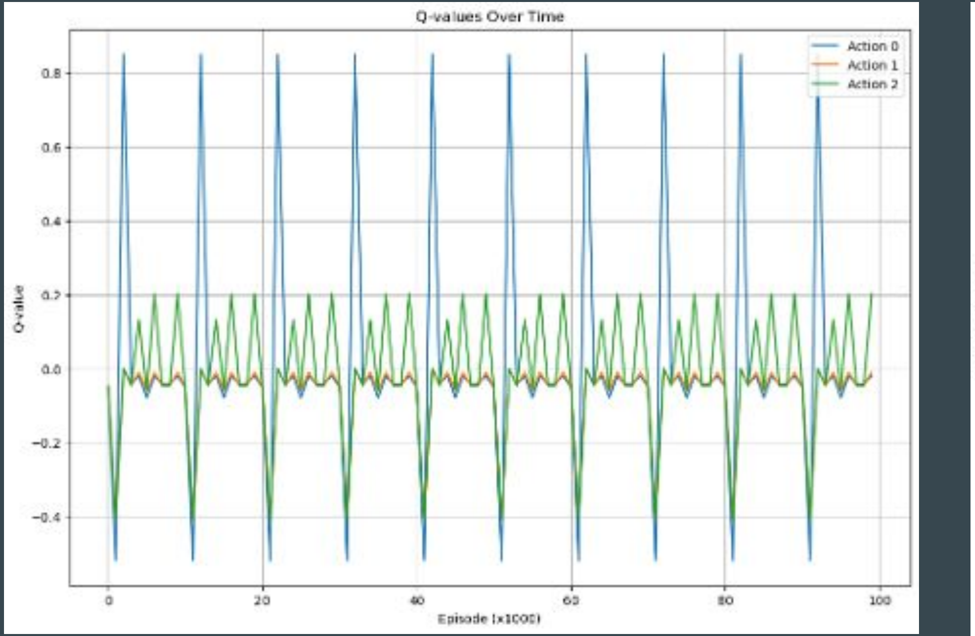
c. Larger penalty

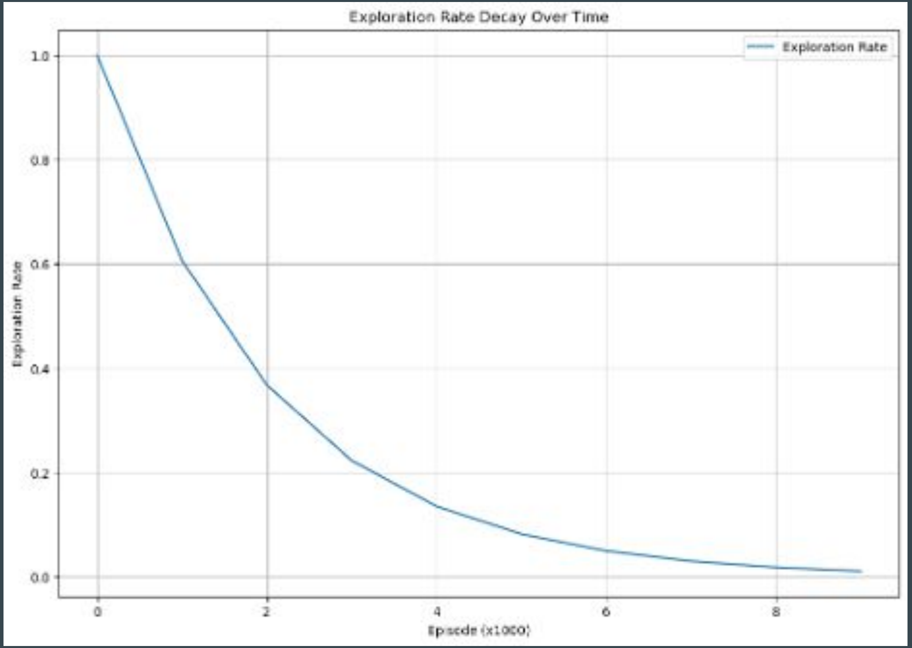




Training

Q-values for three different actions (Action 5%, Action 10%, and Action 15%) over 5,000 episodes of training



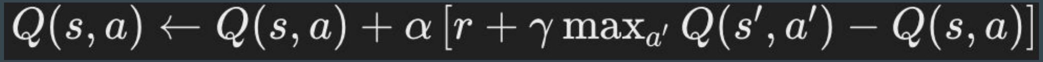


The Math behind:

- **Decision Making**: In Q-learning, the agent typically selects actions by choosing the one with the highest Q-value for the current state. This is based on the principle of maximizing expected rewards.
- **Learning**: During training, Q-values are updated using experiences gained from interactions with the environment. The updates are done using the Bellman equation, which recursively estimates Q-values based on the sum of immediate rewards and the discounted future rewards.

Bellman Equation for Q-values:

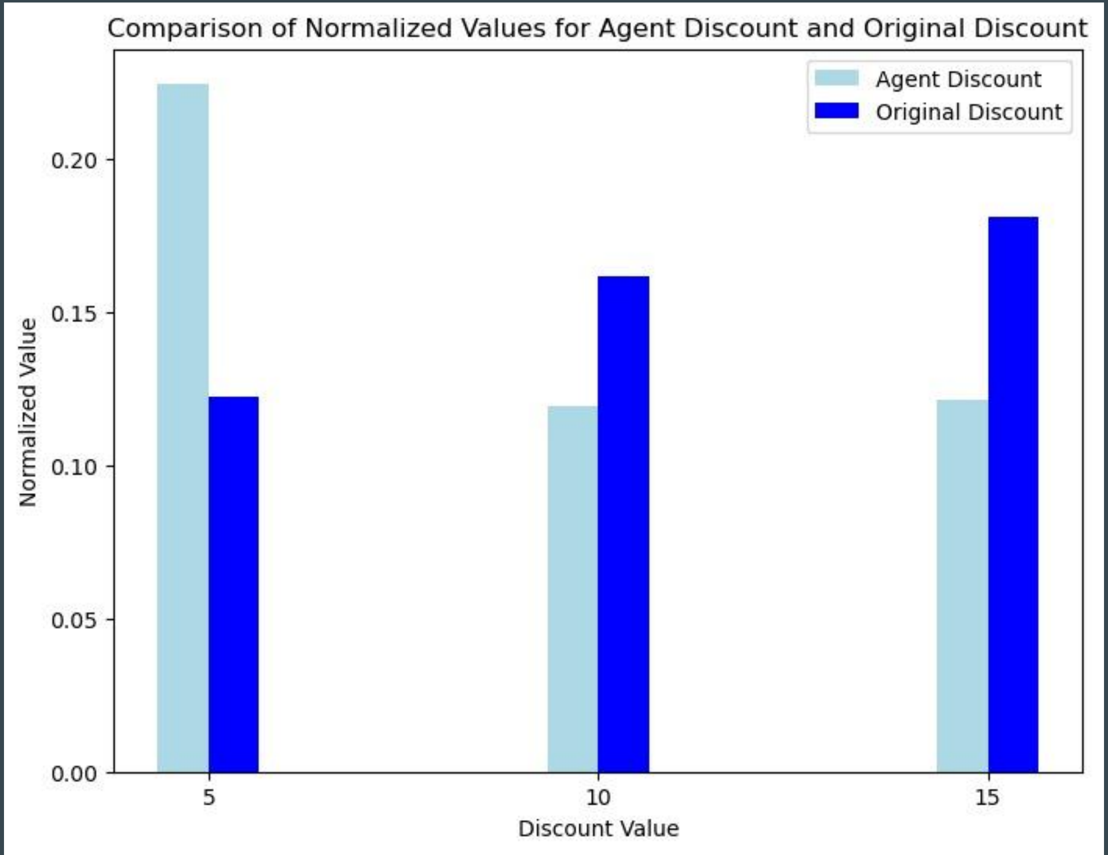
The core of Q-learning is the Bellman equation, used to update the Q-values. It's given by:



$$Q(s,a) \leftarrow Q(s,a) + \alpha \left[r + \gamma \max_{a'} Q(s',a') - Q(s,a)\right]$$

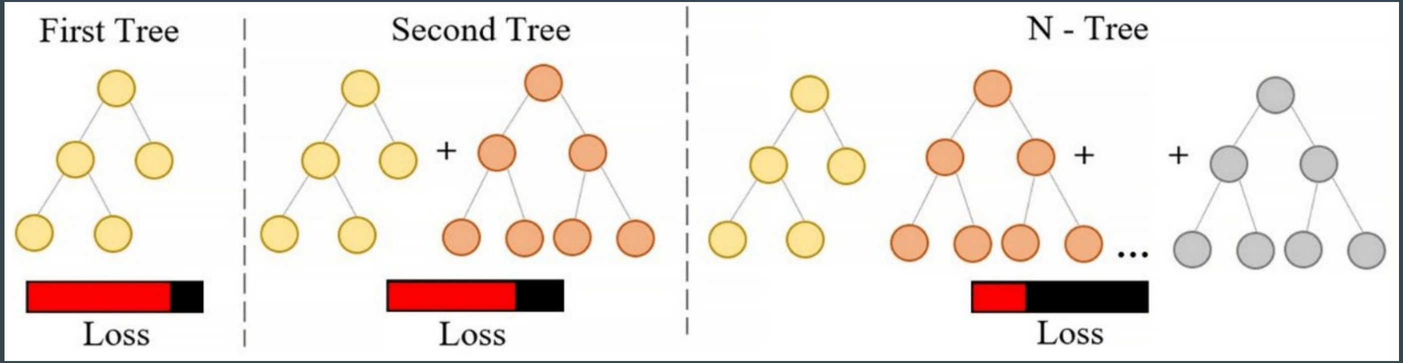
- s′ is the new state after action aaa is taken.
- a′ is a potential future action taken in state s′s's′.
- r is the reward received after taking action aaa in state sss.
- alpha(α) is the learning rate, determining how much new information overrides old information.
- gamma(γ) is the discount factor, which quantifies the difference in importance between future rewards and immediate rewards.

Actions

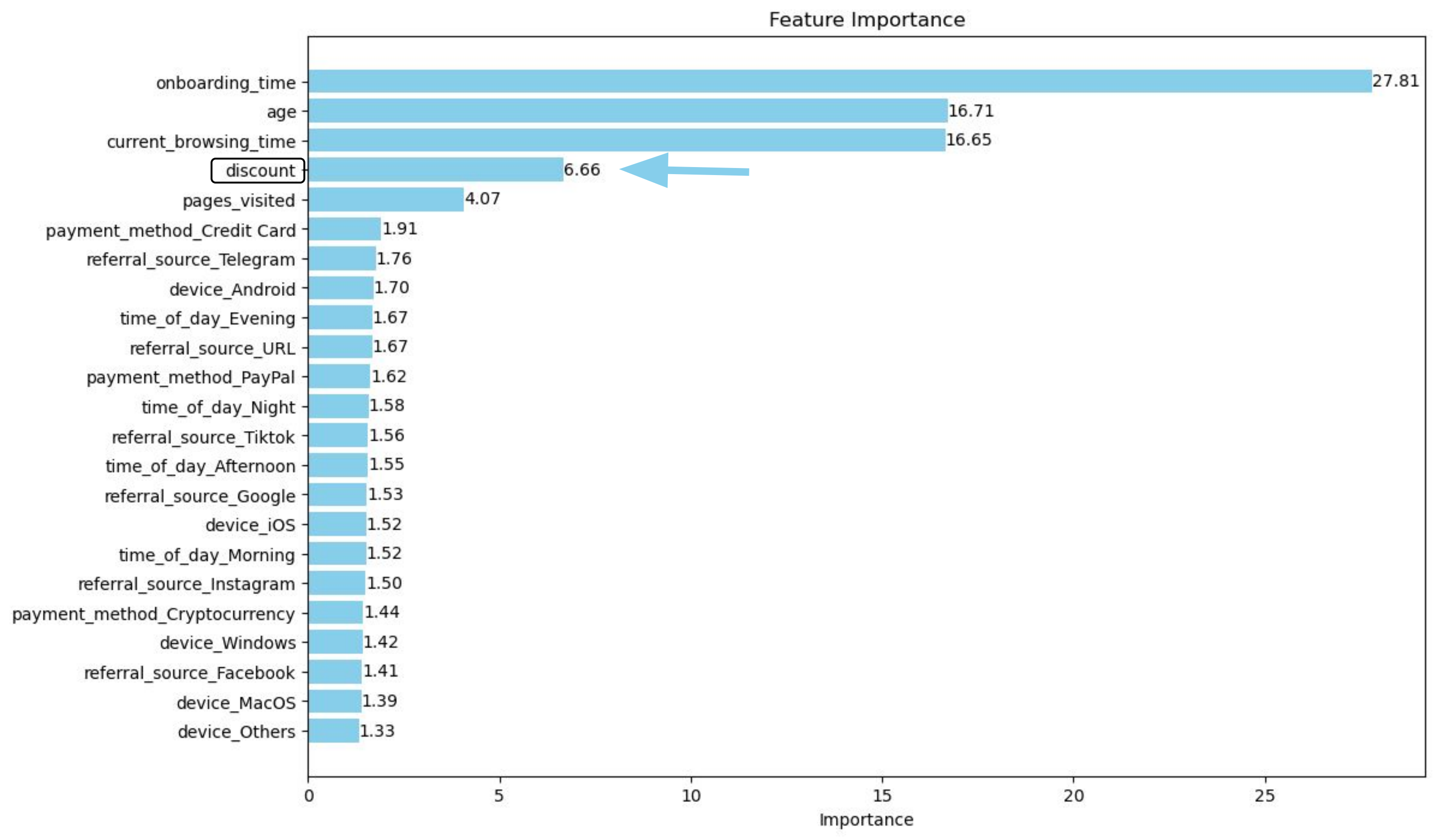

Comparison of Normalized Values for Agent Discount and Original Discount
Agent Discount
Original Discount
Normalized Value
0.00
0.05
0.10
0.15
0.20
5
10
15
Discount Value

Testing

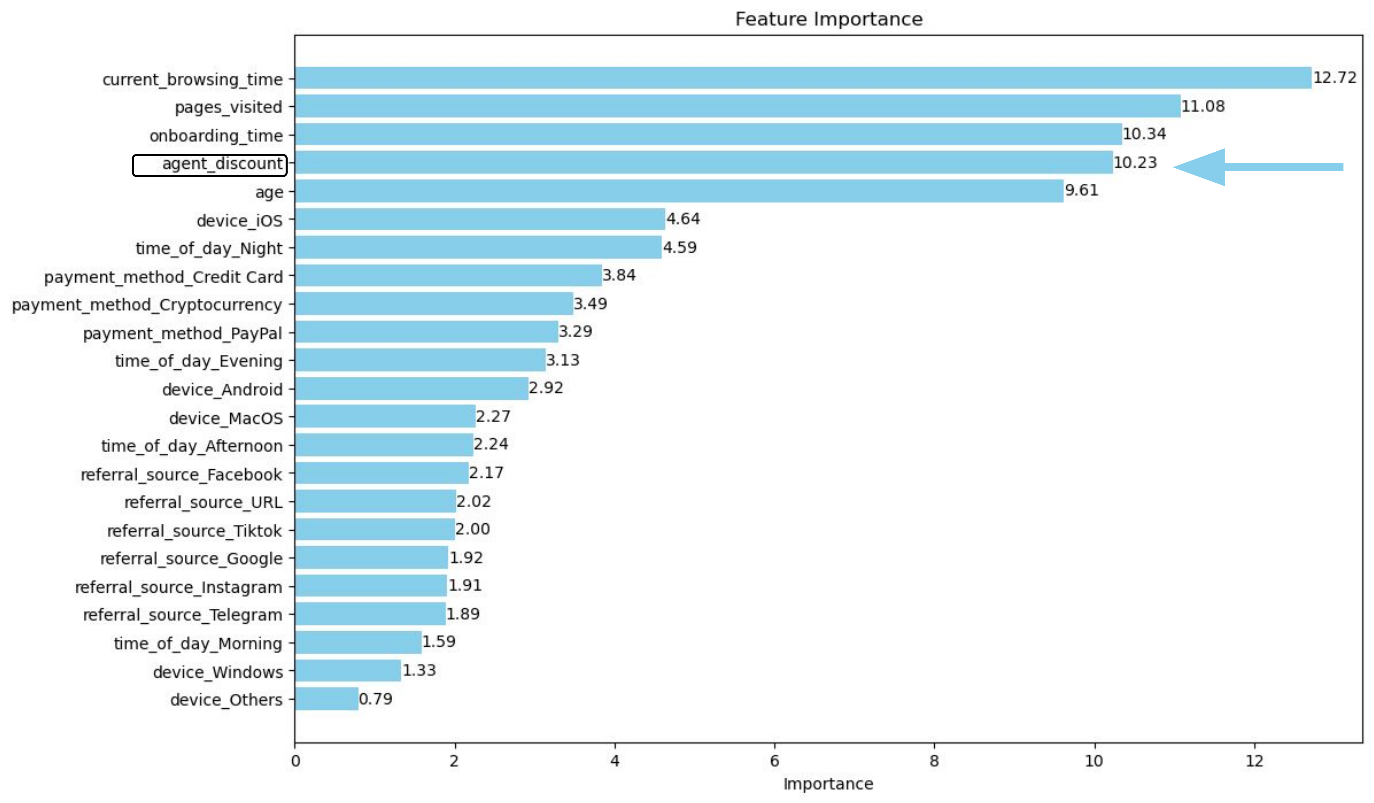
- We used CatBoost feature importance.
- Running it on the initial randomize discount, then on the discount offered by the trained agent.
- We wanted to see how much the discount by the agent is an important feature for the actual deposit.



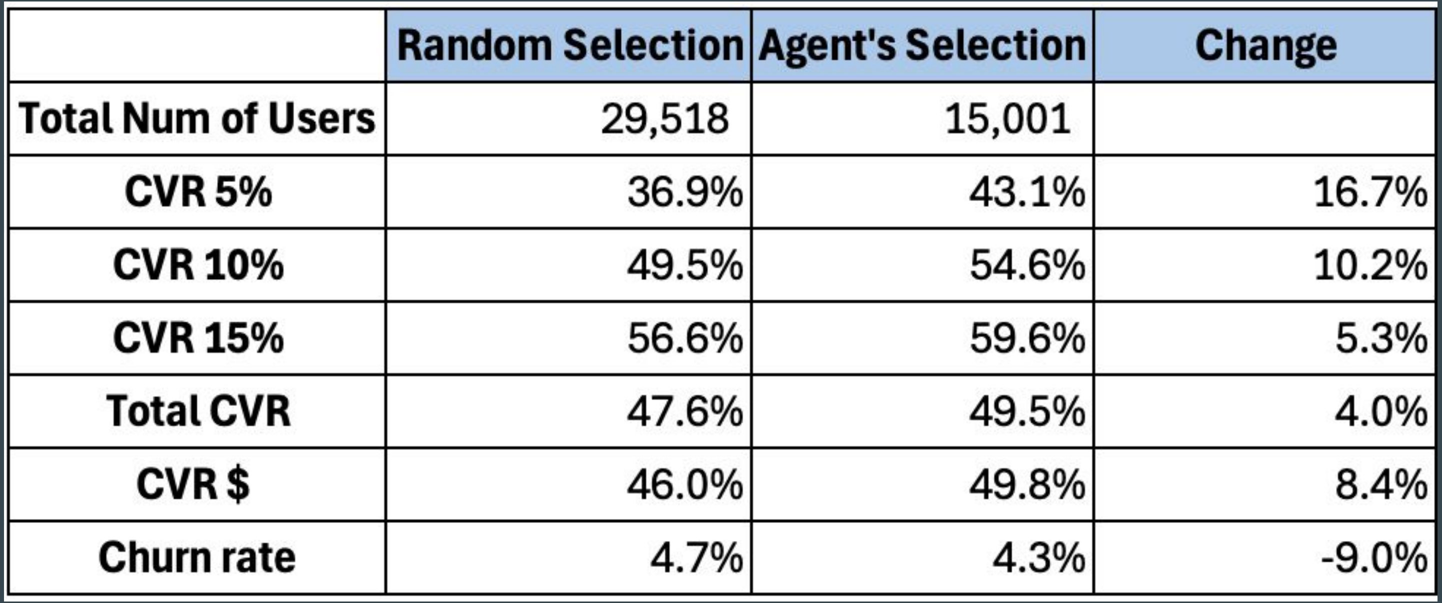
Random Discount



Agent Discount



Results Summary



	Random Selection	Agent's Selection	Change
Total Num of Users	29,518	15,001	
CVR 5%	36.9%	43.1%	16.7%
CVR 10%	49.5%	54.6%	10.2%
CVR 15%	56.6%	59.6%	5.3%
Total CVR	47.6%	49.5%	4.0%
CVR $	46.0%	49.8%	8.4%
Churn rate	4.7%	4.3%	-9.0%

Conclusions

Summary:

- Our agent was proven effective:
    - Increase in approval rates by ±8%
    - Decrease in churn rates by ±9%
- Reinforcement learning provided a powerful framework for advanced A/B testing evaluating several variants in parallel.

Conclusions

Future Work:

- Agent’s implementation on other checkpoints
- Explore more features - we can collect more data during the onboarding.
- Enhance the current model - the training was not stable. Find better hyperparameters
- Deep learning approach - advanced models based on Deep Q-Networks (DQN) might result better results.

Questions?

References

Wang, R., & Blei, D. M. (2019). Dynamic Causal Effects Evaluation in A/B Testing with a Reinforcement Learning Framework. Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 964-974.

https://arxiv.org/abs/2002.01711