Advertisement Data Prediction

Leon Zhang, Ishan Miglani, Qing Shen, Keya Keya

Online Advertisement for Mobile Apps

- Meetsocial, overseas marketing service for Chinese business
- Small-to-medium enterprises, mobile apps
- Advertising strategy advisory based on historical data on previous ads

- A model outputs predicted clicks/installs with features of ad
- Features: spend, channel, os, device, targeting people.
- Goal: cut down unnecessary spend on ads

Meetsocial

Action, Prediction, & Decision-Making

- **The Action** Optimizing advertising settings (targeting ages,genders, budget) for SME campaigns.
- **The Prediction:** Forecasting clicks, installs, and purchases based on campaign settings.
- **The Decision:** Data-backed recommendations to maximize chosen target metric (e.g., maximize clicks within budget constraints).

Data Source

Table 1: 320000 rows with many NAN values, product info, methods, spend, clicks/installs of ad

ad_id	campaign_id	adgroup_id	calendar_date	ucompany_id	product_id	product_name	product_category
15738995501	15738995501	15738995501	2022-01-02	66703	tile.connect.onet.onnect.pairs.matching.game.free.puzzle	Tile Connect - Tile Match Game	
15738995501	15738995501	15738995501	2022-01-03	66703	tile.connect.onet.onnect.pairs.matching.game.free.puzzle	Tile Connect - Tile Match Game	
15712434228	15712434228	15712434228	2022-01-05	32227	com.sevenpirates.idlejp		

campaign_objective	channel	gender	os_type	ad_network_type	device	spend	impressions	clicks	installs	purchase	purchase_value
	Google		ANDROID	SEARCH	TABLET	0	2	0	0	0	0
	Google		ANDROID	SEARCH_PARTNERS	MOBILE	0	2	0	0	0	0
	Google		ANDROID	YOUTUBE_WATCH	MOBILE	30.2455	1874	37	0	0	0

Data Source

Table 2: 30000 rows intersecting table 1 with “ad_id”, providing more information about different ad groups - (targeting age, gender, campaign_objective, etc.)

ad_id	channel	product_id	product_name	product_category	ad_first_dt	ad_last_dt
15712434228	Google	com.sevenpirates.idl	銇勬亼銉砹紒鏀剧彁鎴︹+		2022/1/4	2022/2/16
15738995501	Google	tile.connect.onet.on	Tile Connect - Tile Match Game		2022/1/1	2022/2/8
15745148062	Google	com.meet.andr	Hurrah		2022/1/4	2022/1/4

ucompany_id	campaign_objective	targeting_age	targeting_genders	os_type
32227	APP_INSTALLS	24-65	male	ANDROID
66703	APP_INSTALLS	18-65		ANDROID
16727	APP_INSTALLS	18-65		ANDROID

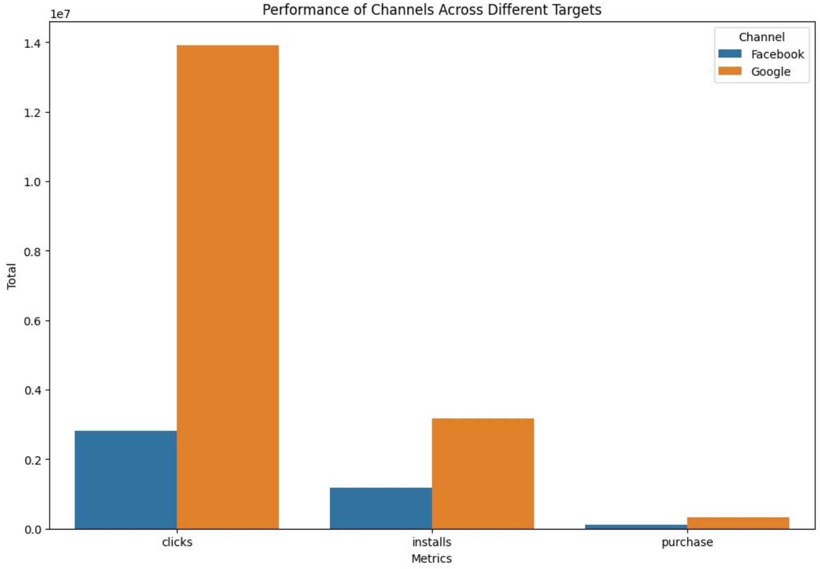
EDA and Feature Engineering

- Huge target variance and skewed distribution
- Merged two tables by their corresponding “ad_id”
- Scraped ratings out of 5, number of ratings and number of reviews
- Built 2 different scrapers for Google Play store and IOS store
- Also scraped categorical information about the apps - whether they are games or productivity apps (eg. categories - PUZZLE, CASUAL, SINGLE PLAYER, STYLIZED, OFFLINE, ROLE PLAYING, BOARD, ABSTRACT STRATEGY, ETC.)

spend	impressions	clicks	installs	purchase	ırchase_valı	PUZZLE	CASUAL	NGLE PLAYE	STYLIZED	OFFLINE	OLE PLAYIN	BOARD	'RACT STRA'
68.9997	4762	73	23	0	0	0	0	0	0	0	0	0	0
30.5187	1722	33	11	0	0	0	0	0	0	0	0	0	0
0.3825	63	3	6	0	0	0	0	0	1	0	0	0	0

Data Preprocessing

- Handling Missing values
- Remove rows with null values
- Create dummy values for categorical features
- Left with ~27K rows and 59 features
- Addressing Skewness: Log Transformations
- Feature Scaling : StandardScaler



Modeling Overview

Model Tested	RMSE	R^2
Linear Regression (Baseline)	1.770	0.478
Random Forest	0.818	0.889
SVR	1.516	0.617
XGBoost	**0.778** ↓ 56%	**0.899** ↑ 88%
Decision Tree Regression	1.064	0.812
Ensemble Model (RF + XGBoost)	0.799	0.894

Target: Clicks

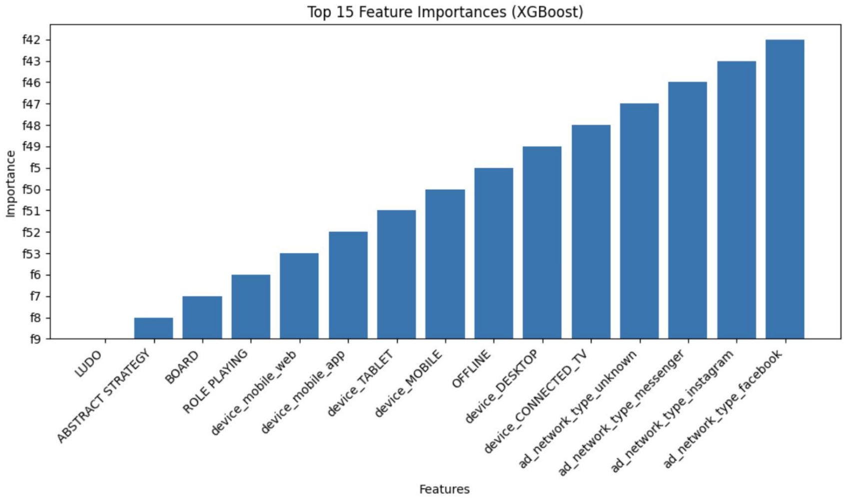
Model selection: Performance metrics of the models - RMSE and R^2

XGBoost Regression Model

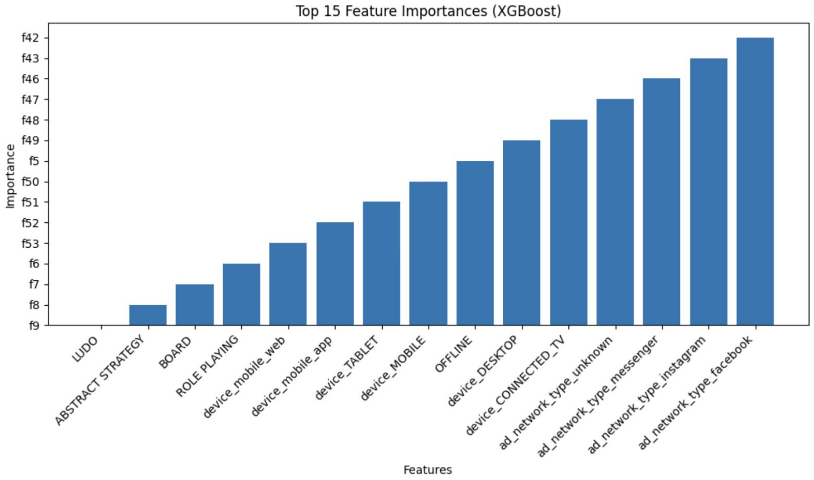
Targets	RMSE	R^2
Clicks	0.778	0.899
Installs	0.775	0.860
Purchase	0.756	0.896
Click/Impression (CTR)	0.869	0.738
Impressions	0.903	0.916

Similar to Clicks, our other targets performed better in XGBoost compared to the other models we tested.

Feature Importance

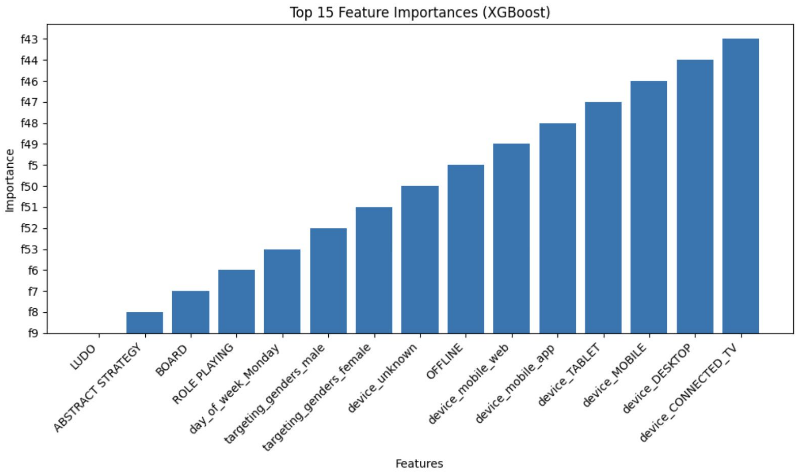


Clicks

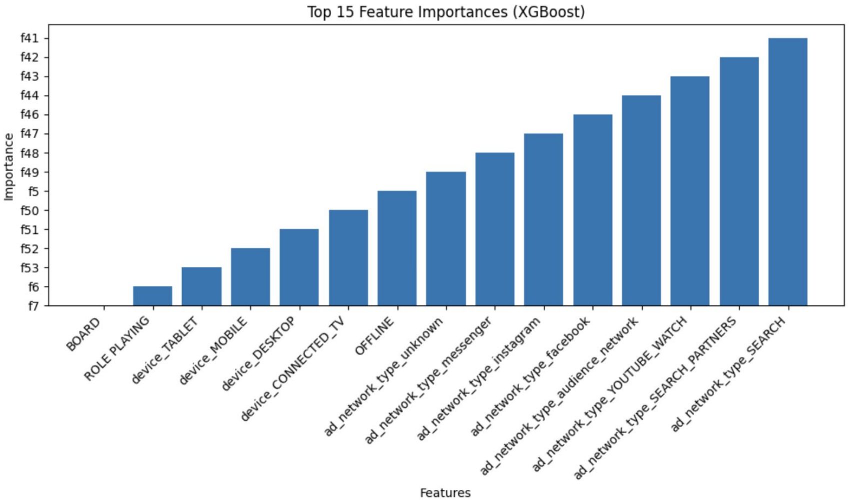


Installs

Feature Importance



Impressions



Purchase

Deployment Plan

Integration with Business Systems: The model will be integrated into existing business systems through APIs, allowing it to access real-time data and provide on-the-fly predictions. This integration ensures that the model's insights are directly usable in the decision-making process.

Cloud Deployment: To handle scalability and ensure accessibility, the model will be deployed on a cloud platform. This approach provides the necessary computational resources and allows for easy scaling as data volume and user base grow.

User Interface: A user-friendly dashboard will be developed to display the model's predictions and insights. This dashboard will allow business users to interact with the model's outputs easily, customize inputs, and visualize data in an understandable format.

Embedding the model in real business scenarios

We generated the 23,040 different combinations by multiplying the number of possible choices for each spend level.

Total Combinations=(Number of Age Combinations)×(Number of Campaign Objectives)×(Number of Channel Google Options)×(Number of OS Type iOS Options)×(Number of Ad Network Types per Channel Google Option)×(Number of Device Types)×(Number of Gender Combinations)

Clubillion Vegas Casino Slots

Triple Sevens: Casino Games

Contains ads · In-app purchases

Social slots casino game offers a collection of completely unique casino slots.

SLOTS | 4.6★ 84.3K reviews | 1M+ Downloads | Teen

Install

Data range: 01-12-2022 to 31-12-2022
Category: Casino
Rating: 4.6255
Num_comments: 12842
Num-rating: 84247

Model's strategy v.s. Human's strategy

With a single ad budget of $2 a day, based on the same budget, the updated ad strategy is expected to result in a 41.9% lift in the number of clicks.

Settings	New Strategy	Previous Strategy
target_age	13-26	45-65
channel	Facebook	Google, Facebook
os_type	Andriod	Andriod
campaign_objective	CONVERSIONS	APP_INSTALLS
ad_network_type	audience_network	All
device	CONNECTED_TV	mobile_app
targeting_genders	female	female,male

spend	predicted_clicks	previous_clicks
2	2.173	1.531
4	2.208	1.676
6	2.251	1.974

Way Forward

Data Collection: Advertiser's Side

- Ad Creative Data:
  - Types and styles of images and videos.
  - Tone and content of text descriptions.
- Advertiser Profile Data:
  - Company size and industry type.
  - Advertising budget levels.
  - Brand reputation metrics.

Data Collection: User Side on the Platform

- User Behavioral Data:
  - Whether a user clicked on an ad link.
  - Duration of ad viewing by the user.
  - Duration of user's stay on the landing page post-click.
  - User interactions with content

Further we can do:

- Richer User Profiles
- Improved Segmentation
- Contextual Relevance
- Competitor Data Analysis

Way Forward

Model evaluation with experts

- Although the model can generate the best advertising strategies, whether these strategies are applicable in real scenarios and whether the recommended parameters are usable still require human evaluation.

Improvements in Predictive Modeling

- Advanced Predictive Models: more complex models like neural networks or transformers
- Hybrid Models: Combine different types of models, such as machine learning models with rule-based systems

Enhancements in Deployment

- Dynamic Feedback Mechanisms: Instead of exhausting all possible combinations in a predictive model, integrate deep reinforcement learning to add dynamic feedback mechanisms