Predicting Purchasing Intent from User Behaviour on e-Commerce Platform

Zaki Hamzah Lee – Data Consultant

Agenda

- Industry Context
- The Problem
- The Data
- Feature Engineering & Selection
- Key Performance Indicators
- Model Training & Validation
- Marketing Strategy

The Industry

USD 4.2 trillion

2021 projected online sales

50% digital

2020-2025 global retail growth

22 billion

June 2020 retail website visits

2.19%

2020 shopper conversion rate

Source: Forbes, Statista

2.19%

2020 shopper conversion rate

What?

Ratio of purchases to visits, expressed as a percentage

So what?

418 million visits across 6 SEA nations in 2020

0.1% conversion rate increase – USD $13.38 million

Who cares?

They do, so we have to

Their 0.1%, our whole business - everyone is a stakeholder

Source: Economic Development Board Singapore, iPrice, SimilarWeb, AppsFlyer

The Problem

- Visit ≠ Purchasing intent
  - Just browsing (Low intent)
- Purchasing intent ≠ Purchase
  - Cheaper competitor (High intent, but alternative cheaper)
  - High checkout fees (High intent, but final price increased)
- User feelings unknown, but user activity known

Business question: Can we identify users with purchasing intent from their user session activity to apply targeted marketing strategies and increase purchases?

Data question: What features and model best predict user purchasing intent from user session data?

The Data

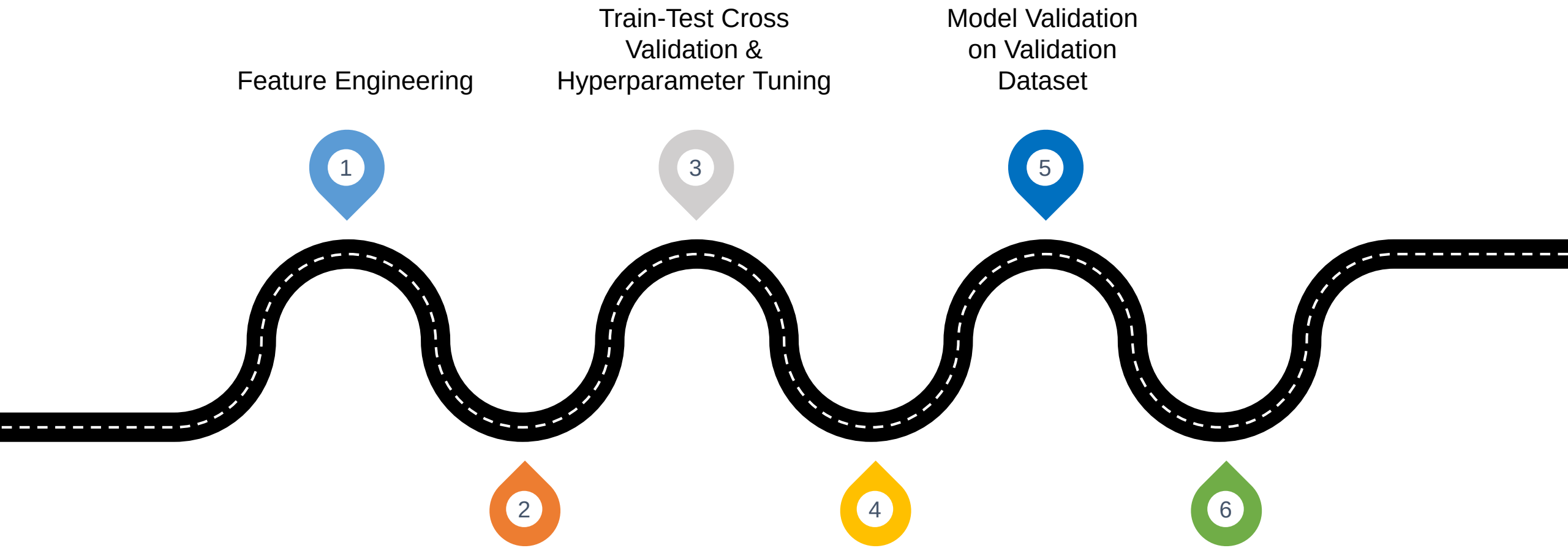
Oct 2020 event activity for Electronics e-Commerce store in Singapore (assumed)

- 161,544 records, 9 variables

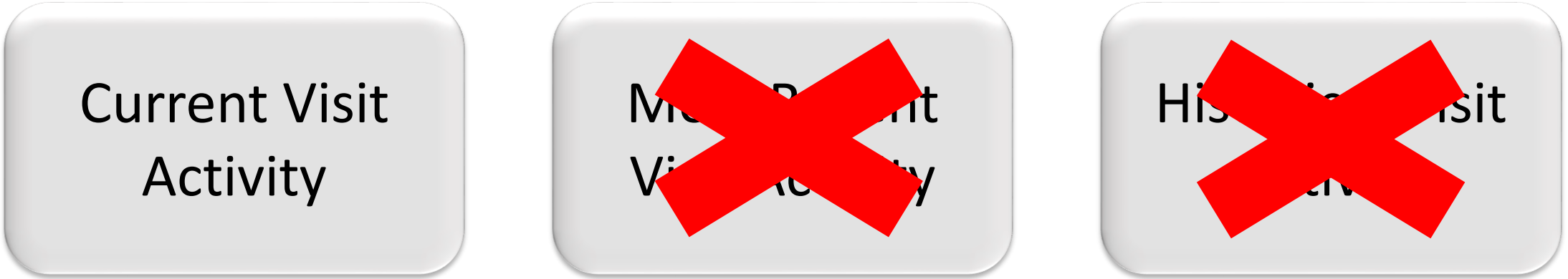
Event Time	Event Type	Product ID	Category ID	Category Code	Brand	Price	User ID	User Session
2020-10-01 00:33:34+00:00	purchase	1168166	2144415927049912542	electronics.video.tv	starwind	105.87	1515915625521281104	SJUH0TBLGD

- Event Types: view, cart, purchase
- “User Session” can be open for days – unreliable
- Daily user activity used as visits
  - 92,277 visits – 3,983 purchases; 88,294 no purchases

Workflow


Feature Engineering
Train-Test Cross Validation & Hyperparameter Tuning
Model Validation on Validation Dataset
1
2
3
4
5
6
80-20 Train-Test Validation Split
KPI Comparison & Model Selection
Final Performance Evalulation

Feature Engineering & Selection



- Most Recent and Historical Visit Activity metrics biased against new users
  - Feature engineering limited to Current Visit Activity
  - Past Visit Activity useful for other applications e.g. customer profiling, rewards program

Feature Engineering & Selection

Feature engineering based on inferred interest

- Time investment
- Repeated product engagement (‘view’ and ‘cart’)
- Pre-purchase action (‘cart’)

Event Time	Event Type	Product ID	Category ID	Category Code	Brand	Price	User ID	User Session
2020-10-01 00:33:34+00:00	purchase	1168166	2144415927049912542	electronics.video.tv	starwind	105.87	1515915625521281104	SJUH0TBLGD

Daily User Session (Index)	Duration in Seconds	Count of ‘Views’ and ‘Carts’ for most frequently browsed Category ID	Cart Quantity	Purchased (Target)	Purchase Value	Cart Value
1515915625521281104 _2020-10-01	1560	4	1	1	105.87	105.87

Key Performance Indicators

- High Recall score – Accurate prediction of purchasing intent
- Model training time

Note: False Positives not model failure

- Predicted purchasing intent, but outcome was no purchase
- Remember: Purchasing intent ≠ Purchase
- Essential target group for marketing strategy

Train-Test Results

- All models had practical training time
- XGBoost highest in Recall score
- Hyperparameter tuning boosted Recall score to 92.9%
  → Selected for validation

Algorithm	Mean Train-Test Recall	Mean Model Train Time (seconds)
Logistic Regression	28.2%	0.332
Support Vector Machine	33.5%	47.8
Naïve Bayes	60.2%	0.0195
Decision Tree	60.9%	0.0419
Random Forest	65.5%	1.91
AdaBoost	80.7%	0.965
XGBoost	89.2%	1.60
XGBoost (tuned)	92.9%	

Validation

- Recall score: 93.0%

→ Highly accurate purchasing intent prediction

→ All 56 FN were purchases without ‘cart’ event

- Feature Importance (normalized):

1. Cart Quantity (0.895)
2. Count of ‘Views’ and ‘Carts’ for most frequently browsed Category ID (0.104)
3. Duration in Seconds (0.001)

Predicted Outcome

Actual Outcome

	No Purchase	**Purchase**
No Purchase	17,280 (TN)	379 (FP)
Purchase	56 (FN)	741 (TP)

Marketing Strategy

- Adding items to cart key indicator of purchasing intent

Focus strategy on top causes of cart abandonment:

1. Just browsing/Not ready to buy – 58.6%
2. Extra costs too high (shipping, tax, fees) – 20.3%

Predicted Outcome

Actual Outcome

	No Purchase	**Purchase**
No Purchase	0.01% Cart Quantity >0	100% Cart Quantity >0
Purchase	0% Cart Quantity >0	100% Cart Quantity >0

Source: Baymard Institute

Marketing Strategy

Focus area: Cart abandonment due to extra costs too high – 20.3%

Focus Segment

→ 20.3% * (242 + 379) = 126 additional purchases

Cost: Apply average $5 discount – taken up by focus segment + confirmed purchases

→ $5 * (126 + 741) = $4,335

Average cart amount = $141

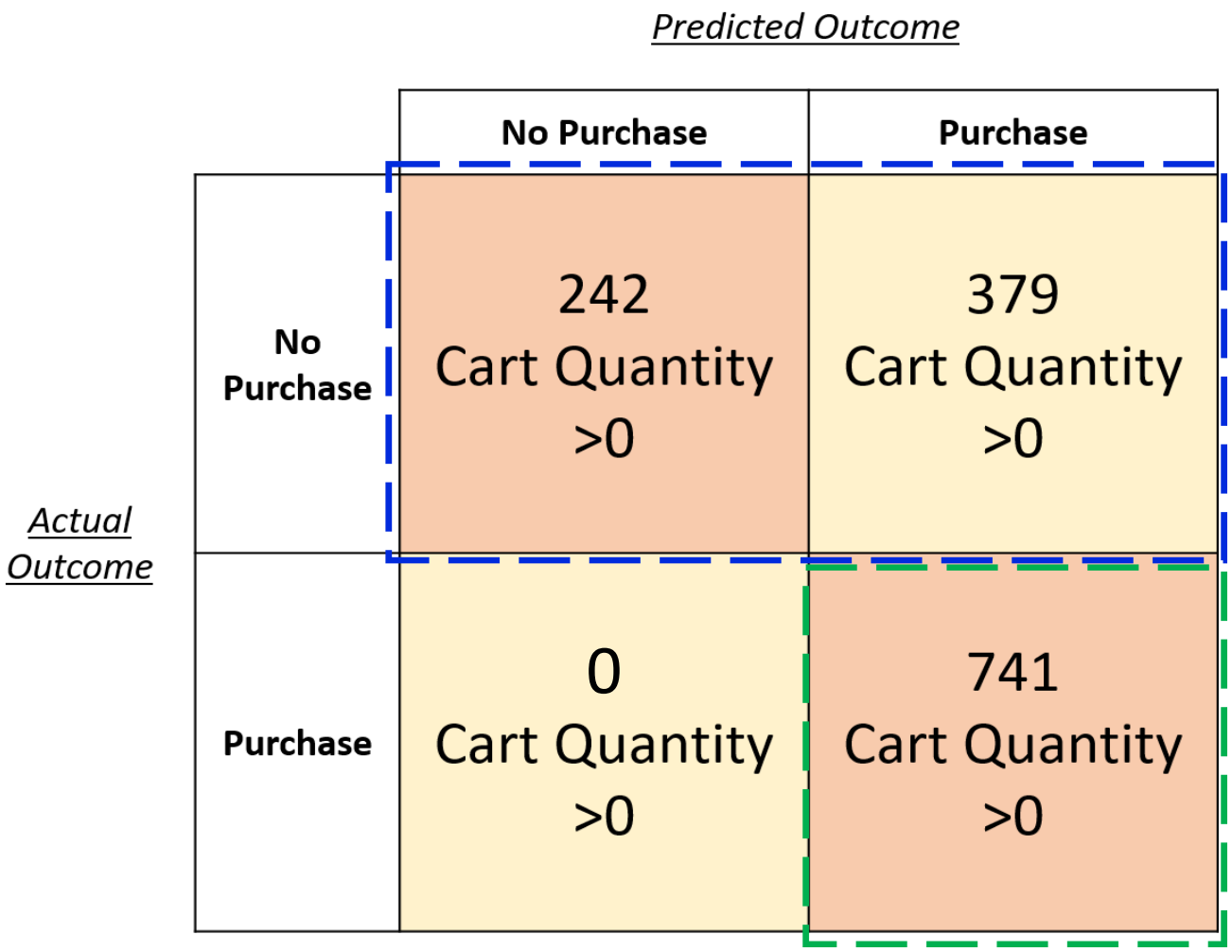
Additional revenue

→ 126 * $141 = $17,766

Assume 5% of monthly revenue for marketing budget

→ 5% * $562,590 = $28,129

Strategy is within budget and brings more revenue than cost



Conclusion

1. *Adding items to cart is a key indicator of purchasing intent*
2. *Tuned XGBoost model attained 93.0% recall score for predicting purchasing intent*
3. *Marketing strategies can be applied to users with predicted purchasing intent to prevent cart abandonment and increase conversion rate*

Future steps

- Validate marketing strategy with trial
- User activity can be enriched with more variables

Thank You

Appendix

Data Source:

https://www.kaggle.com/mkechinov/ecommerce-purchase-history-from-electronics-store

Forbes:

https://www.forbes.com/sites/joanverdon/2021/04/27/global-ecommerce-sales-to-hit-42-trillion-as-online-surge-continues-adobe-reports/?sh=14404f7150fd

https://www.forbes.com/sites/michelleevans1/2021/03/25/global-e-commerce-market-to-expand-by-us1-trillion-by-2025/?sh=5f3df336cc02

Statista:

https://www.statista.com/statistics/1112595/covid-19-impact-retail-e-commerce-site-traffic-global/

https://www.statista.com/statistics/439576/online-shopper-conversion-rate-worldwide/

Appendix

Google Analytics:

https://support.google.com/analytics/answer/6014873?hl=en#zippy=%2Cin-this-article

Economic Development Board Singapore, iPrice, SimilarWeb, AppsFlyer:

https://www.edb.gov.sg/en/business-insights/insights/how-did-the-pandemic-affect-online-shopping-behaviour-in-2020.html

Baymard Institute:

https://baymard.com/lists/cart-abandonment-rate