Quadrangle Club

Jyoti • Eddie • Zhe

1998
2017
1998: "I sell books."
2017: "I sell whatever the f I want."

Clarify Target

Target

Given user demographic & app interaction data

Predict probability of purchase

Difficulty

Decrease data size

Create label

Feature engineering

Workflow

Individually processing

↓

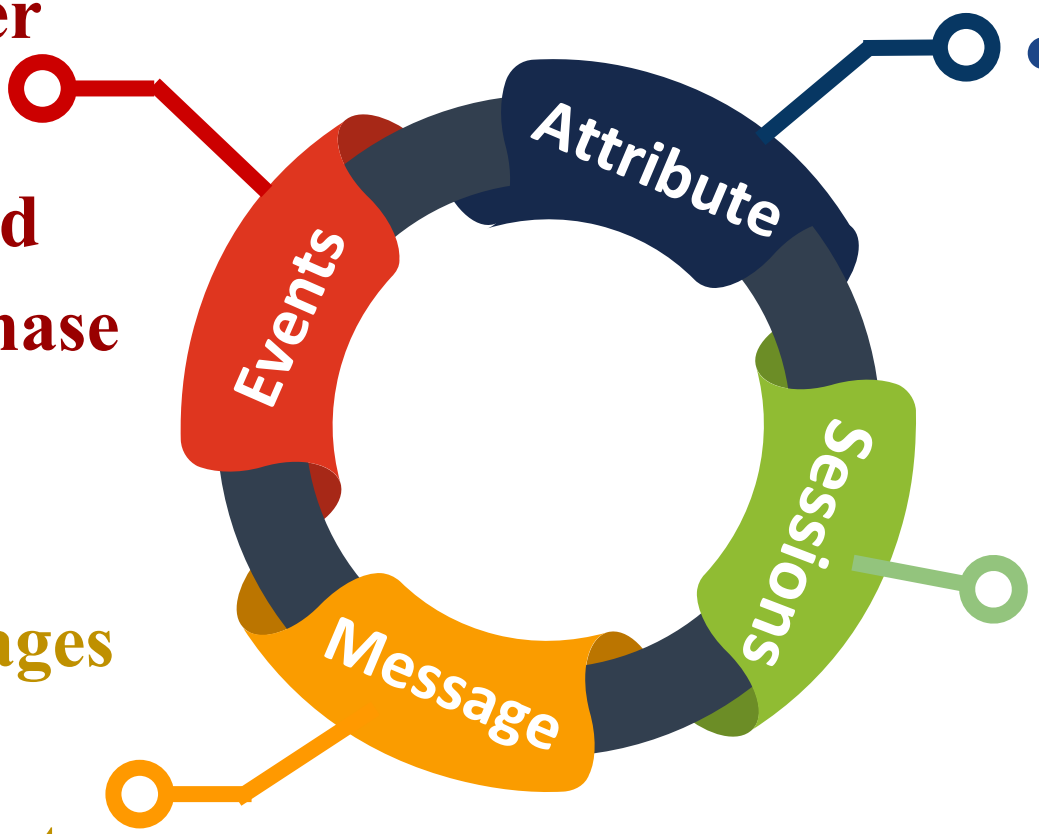
Combination

↓

Improve

Understand Data

- Each row represents one “event” one user created
- All columns encoded
- Event code 8: purchase



- User profile data
- Provide plenty of information to draw portraits of user

- Time-series dataset
- Track user activity

- Meta-data on messages sent to user
- Directly link to Events data

Feature Engineering

Create Label

- Dec 1st - Dec 14th for label
- Previous two months data for training

Sessions

- Aggregate features
- Different time intervals
- Capture the time series data

Messages

- merged to other data

Events

- 90 unique event codes
- Compute the total counts and sum of values

Attributes

- Label encoding for user id
- Use pivot_table
- Use regular expression to parse attribute 66

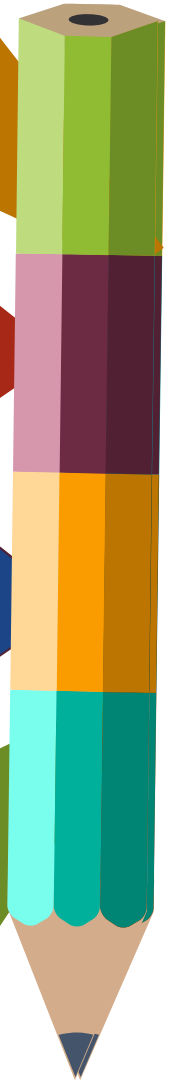
Features Menu

Messages — Sold Out

Events — 'user_id_hash', '7_y_4', '45_x_4', '8_y_4', '8_x_4', '7_x_4', '1_x_4', '5_x_4', '5_y_4', '14_x_4', '6_x_4', '4_x_4', '40_x_4', '42_x_4','3_x_4', '41_x_4', '45_x_3'.....

Attributes — 1_total_7d, 15, 1_total, 15_total, last_day_session_count, 15_total14d, 14, 8_y_4, 1_total_14d, 8_x_4, 19, 7_y_4, 15_total7d ……

Sessions — user_id_hash, session_count, last_7_d_session_count, last_day_session_count, last_14d_session_count, device_count, prev_sess_mean prev_sess_median……



Model Performance

Play with single dataset

Sessions	0.9260
Attributes	0.9882
Events	0.9205

Combine different datasets

Sessions	
Attributes	0.9906

Welcome Gradient Boosting

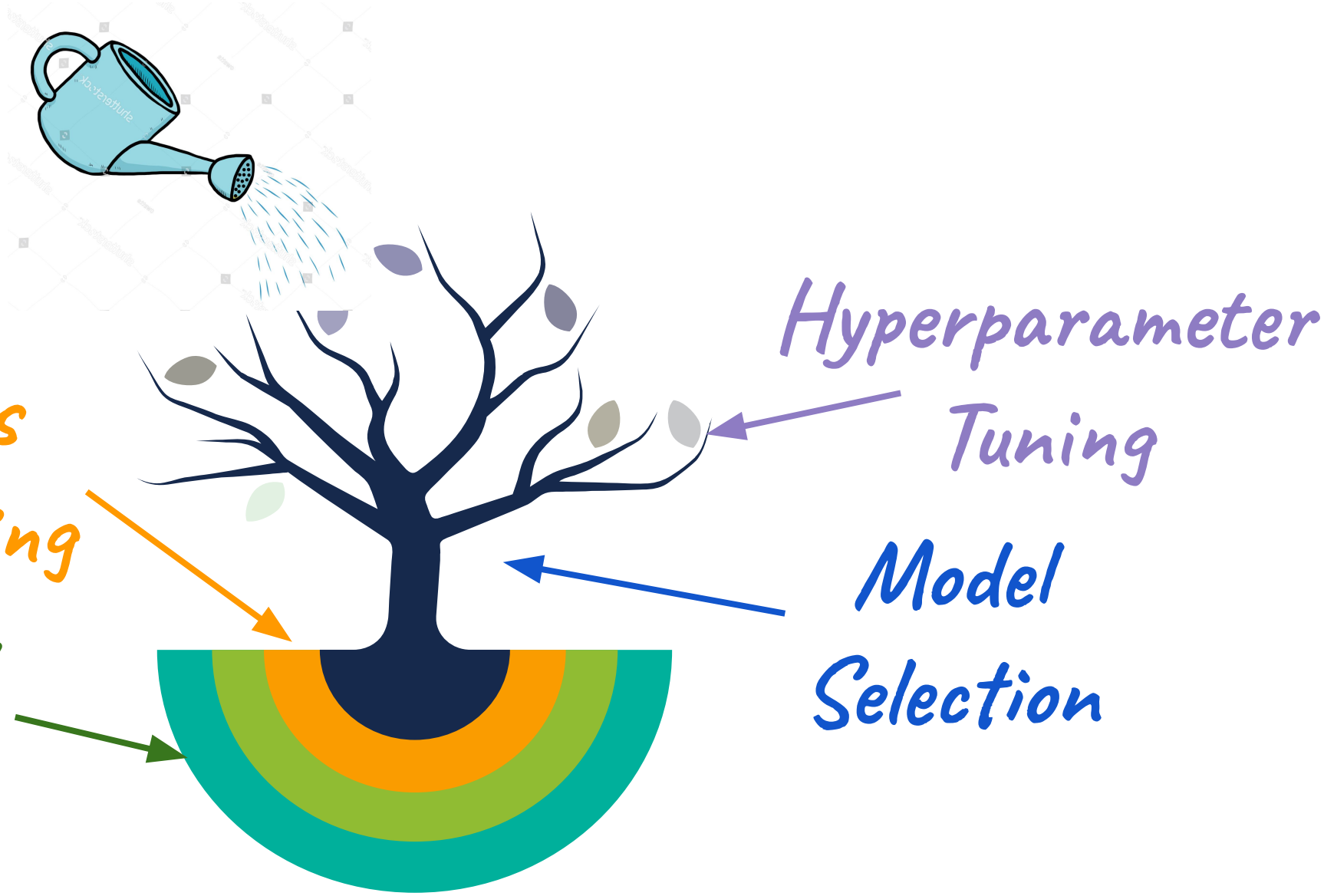
Sessions	
Attributes	0.9929

Let's put things together

Final AUC	0.9951

Lesson Learned

Team Cooperation !!



THANKS

To Yannet
To all our great classmates
To our team members
To Leanplum and Kaggle