SI 671 Final Project: E-commerce User Behavior Analysis and Modeling

Xinyu Zhang (xyuzhang@umich.edu)

OVERVIEW

Every business decision takes much financial and human resource costs. It is important to understand customers' interests and behavior to drive sales. This project aims to conduct user behavior analysis for e-commerce platforms to gain insights and make effective business decisions.

Research Questions:

- What are the peak activity times in purchasing?
- Are top-selling categories the ones that receive the most views?
- Can we predict when a user is likely to stop using the platform or reduce their engagement?
- What is the next interaction of a user towards one item?

DATASET

The dataset used is sampled from ***User Behavior Data from Taobao for Recommendation.***

- Records of one million users between November 25, 2017 and December 3, 2017
- Four behaviors: page views, purchases, adding to the cart and favoriting products
- **Data Prerocessing**: data sampling, cleaning, format transformation …

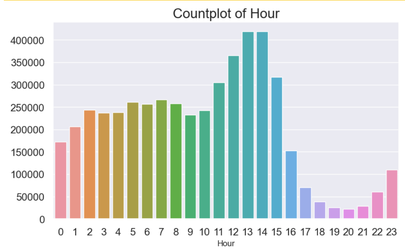
	User_ID	Product_ID	Category_ID	Behavior	Timestamp
1747	1000060	2178355	4962280	pv	1511642462
1748	1000060	857323	1320293	pv	1511822512
1749	1000060	4983347	4069500	pv	1511822555

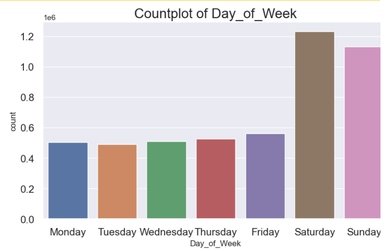
Original Dataset: 100150807 records with 987994 unique users

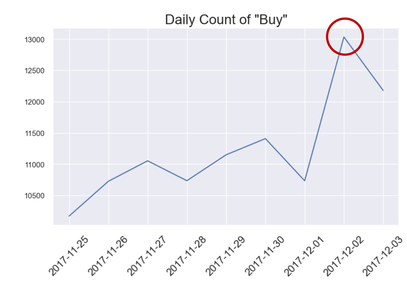
	Datetime	Day_of_Week	Hour	Date
1747	2017-11-25 20:41:02	Saturday	20	2017-11-25
1748	2017-11-27 22:41:52	Monday	22	2017-11-27
1749	2017-11-27 22:42:35	Monday	22	2017-11-27

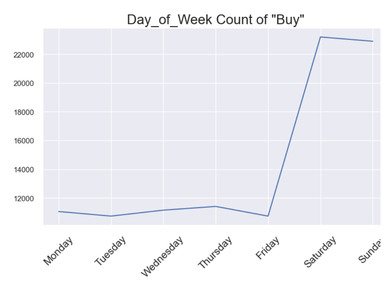
Processed Dataset: 4952632 records with 49400 unique users

DATA ANALYSIS







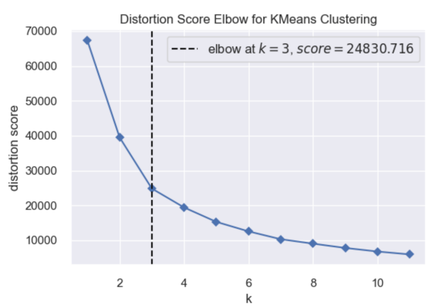


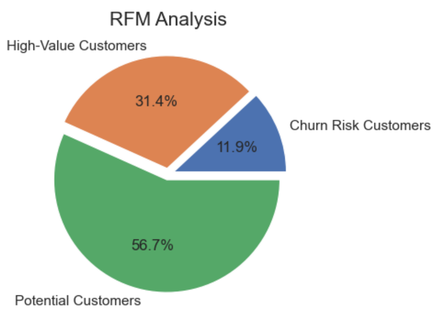
CUSTOMER SEGMENTATION

RFM Analysis with K-means Clustering

- **RFM Analysis** uses past purchase behavior to divide customers into groups for future marketing strategy
  Recency: days since last purchase
  Frequency: total number of purchases
- **K-means Clustering with Recency and Frequency**
  Use **Elbow method** to select the optimum k value

❑ Churn Risk Customers: high recency + low frequency
❑ Potential Customers: moderate recency
❑ High-Value Customers: low recency + high frequency

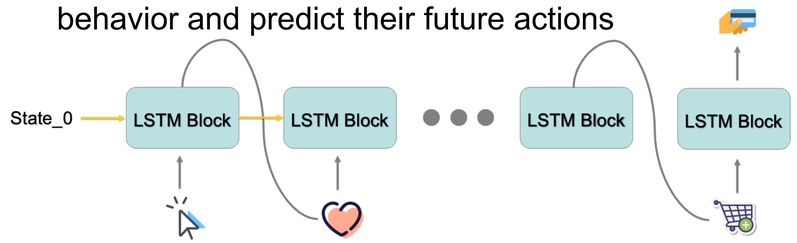




BEHAVIOR MODELING

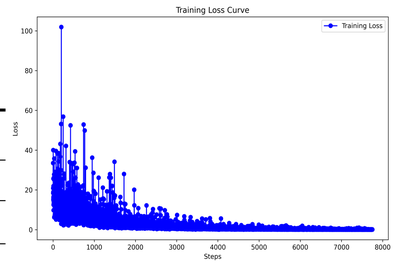
Method Overview

- **Long Short-Term Memory (LSTM)** is a variant of recurrent neural network (RNN), using memory cell structure to capture sequential dependencies in user behavior and predict their future actions



Experimental Results

Methods	Test Accuracy
Random Choice	0.25
RNN	0.73
LSTM	0.95



CONCLUSION & DISCUSSION

- Users interests of purchasing greatly increase during weekends and shopping festivals
- The conversion rates are exceedingly low, mostly below 4% even among the top-selling categories
- 11.9% of users are churn-risk users but the overall customer groups contribute positively to Taobao
- LSTM model predicts the next user behavior with an accuracy of 95%, showcasing its feasibility

FUTURE

- Use frameworks like Spark or computing resources like GPU to process the whole dataset
- LSTM models can be further applied with recommendation systems to predict the user preferences of different products and categories