

# What You'll Buy Next

---

# & When You'll Buy It

Utilizing AWS and Spark ML to Predict User Choices

**Team 9:** Xiyang Xu, Kitlyn (Hui-Lun) Kuo , Yanlun Ren, Raghuveer Rao Vijjini, Sydney Jiang, Wade Wimer

# Agenda

A decorative blue wavy graphic that flows across the top of the slide, starting from the left and tapering off towards the right.

**Predictive: System Concept and Comparison to Traditional**

**Tools and Emergent Techniques**

**System Creation and Output**

**Business Use Cases**

**Conclusion**

# Agenda

A decorative graphic consisting of several overlapping, wavy blue lines that create a sense of motion and depth, extending across the top of the slide.

**Predictive: System Concept and Comparison to Traditional**

**Tools and Emergent Techniques**

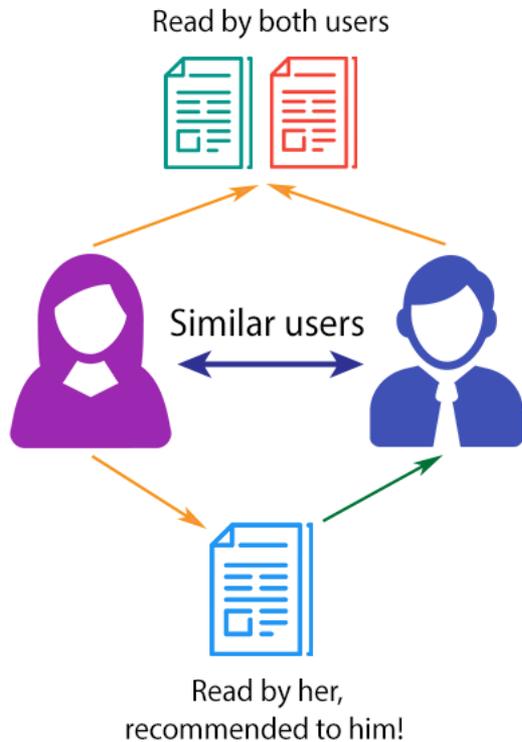
**System Creation and Output**

**Business Use Cases**

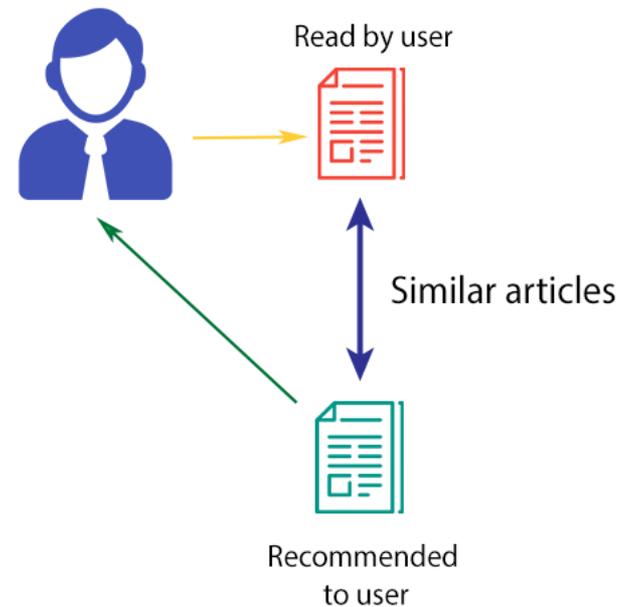
**Conclusion**

# Recommender System Concept

## COLLABORATIVE FILTERING



## CONTENT-BASED FILTERING



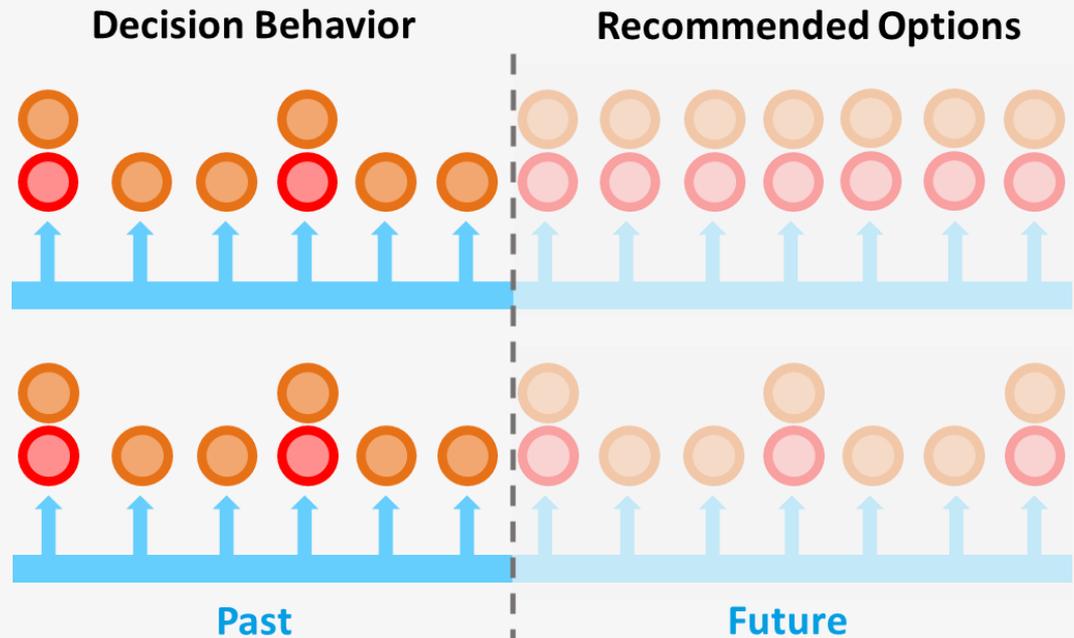
# Comparison to Traditional Systems

## Traditional Systems:

Suggest products that customers want, but do not address when they may want them.

## “Predictive” System:

Sophisticated models introduce time as a predictive factor, providing helpful reminders, not ads



# Agenda

A decorative blue wavy graphic that flows across the top of the slide, starting from the left and curving downwards towards the right.

**Predictive: System Concept and Comparison to Traditional**

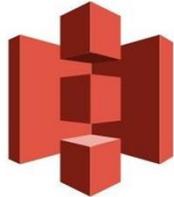
**Tools and Emergent Techniques**

**System Creation and Output**

**Business Use Cases**

**Conclusion**

# Tools and Emergent Techniques



Amazon S3

**Recommender systems** use large amounts of proprietary data gathered for several years

**S3 provides:**

- Unlimited Storage
- Immediate Access
- Security



Amazon EMR

**Recommender systems** require robust computational strength to manipulate high volume data with complex algorithms

**Amazon EMR's** computing structure is streamlined for big data capabilities



Modern **Recommender systems** are complex and intentional algorithms with many stages

**Spark ML's** algorithms interface with big data applications and allow for recommender system creation for free with few key strokes



# Agenda

A decorative blue wavy graphic that flows across the top of the slide, starting from the left and tapering off towards the right.

**Predictive: System Concept and Comparison to Traditional**

**Tools and Emergent Techniques**

**System Creation and Output**

**Business Use Cases**

**Conclusion**

# System Creation



**Input Data**

**Feature  
Engineering**

**Model  
Building**

**Output  
Translation**



# System Creation: Input Data

User ID

Order Number

Transaction Time

Product

Amt Purchased

Last Visit Time

Item Location



Minimum attributes needed  
to inform model

Input Data

Feature  
Engineering

Model  
Building

Output  
Translation

# System Creation: Feature Engineering

Input Data

Feature Engineering

## Developed Sections

User Features

6 Features

Item Features

9 Features

User x Item Features

7 Features

Datetime Features

2 Features

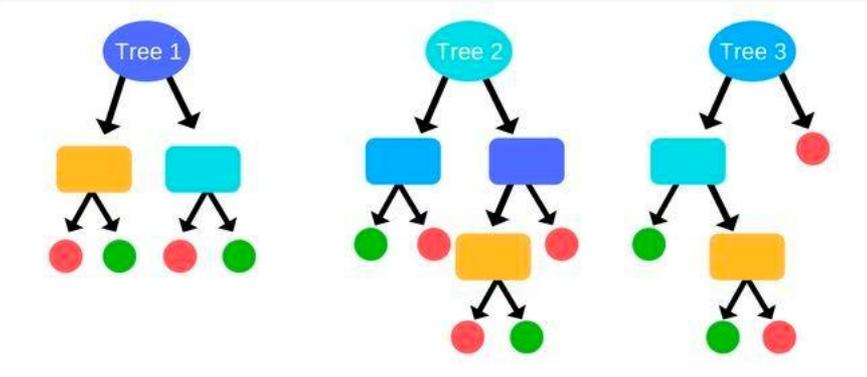
Input Data

Feature Engineering

Model Building

Output Translation

# System Creation: Model Building



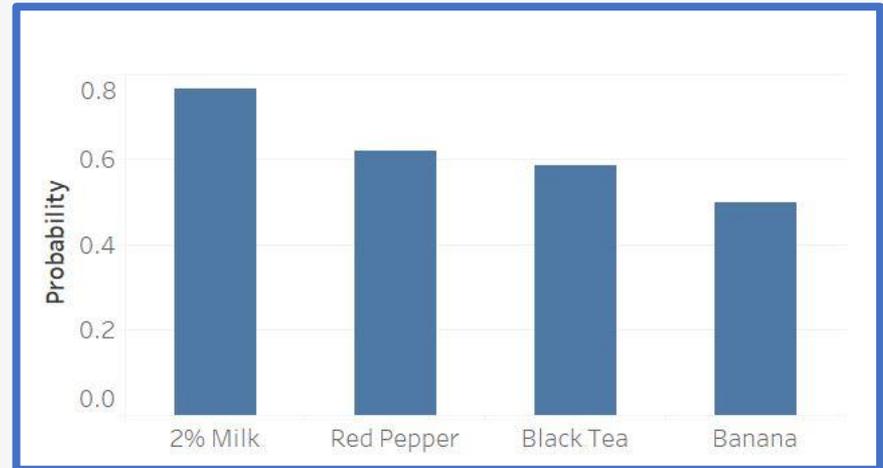
- ★ Random Forest Algorithm
- ★ Good performance

- ★ 50 tree nodes running parallel across Spark clusters



# System Creation: Output Translation

order_id	product_id	pred
1054066	11520	0.12956429
1054066	14211	0.18618035
1054066	35461	0.05001443
1054066	3339	0.06505957
1054066	13452	0.11622027
1054066	24852	0.08288204
1054066	27288	0.13649434
1054066	40604	0.07771818
1054066	17706	0.17063484
1054066	47156	0.42731799
1054066	3896	0.17063484
1054066	16953	0.06505957
1054066	4792	0.08031924
1054066	36929	0.09259171
1054066	27086	0.12956429
1054066	21070	0.13125784
1054066	48974	0.05232617
1054066	27729	0.08752166
1054066	49235	0.08288204



probabilities of a specific user buying specific product at a specific time

Input Data

Feature  
Engineering

Model  
Building

Output  
Translation

# Model Case Example



## User: Sydney

- Healthy purchases on Monday afternoon
- Snacks on Weekend nights

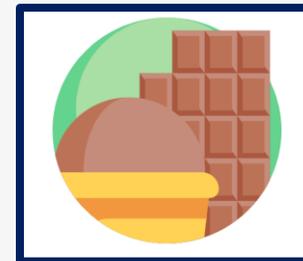
### Monday Afternoon Basket

- Whole wheat bread
- Tomatoes
- Coconut Milk
- Lemon Juice



### Sunday Evening Basket

- Chocolate nuts
- Chocolate Mint Ice Cream
- Tortilla Chips
- Case Sugar



# Agenda

A decorative graphic consisting of several overlapping, wavy blue lines that create a sense of motion and depth, positioned at the top of the slide.

**Predictive: System Concept and Comparison to Traditional**

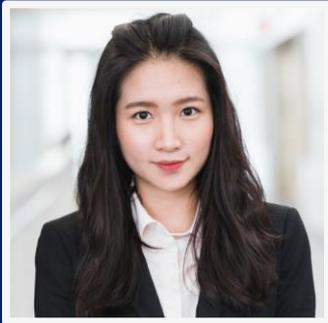
**Tools and Emergent Techniques**

**System Creation and Output**

**Business Use Cases**

**Conclusion**

# E-Commerce Use Case



## User: Kitlyn

- Browse app weekly at 8pm
- Prefer healthy products

### Purchased Item



- Toothpaste
- Repurchase frequency: monthly
- 90% repurchase rate

### User x Item Features



- 30 days since last toothpaste order
- Co-purchase with mouthwash

### Recommender System:



**Recommend toothpaste and mouthwash a month from her last purchase.**

# Financial Service App Use Case



## User: Wade

- Trades stock once per week
- Browses app daily 6pm
- Prefers tech

### Purchased Item



- APPL stock
- High growth potential
- Technology sector

### User x Item features



- Repurchase stock in 10 days
- 30% of user's portfolio

### Recommender System:



**Recommend purchase MSFT in 10 days from repurchase AAPL stock.**

# Other Use Case Examples



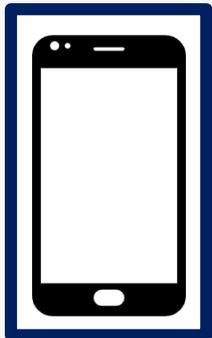
## Healthcare

- Diet
- Drugs
- Checkups



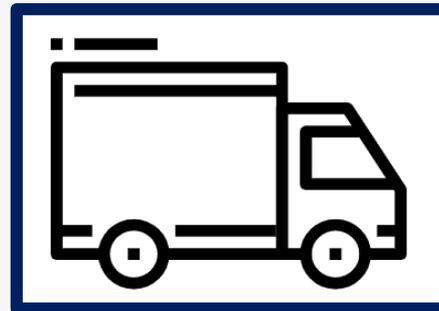
## Taxi / Rideshare

- Driver Insights
- Consumer Reminders
- Car Maintenance



## Smartphones

- Notifications
- News
- Online Videos
- Apps/Websites



## Shipping / Logistics

- Regular Shipments
- Scheduling and Directions

# Agenda

A decorative graphic consisting of several overlapping, wavy blue lines that create a sense of motion and depth, positioned at the top of the slide.

**Predictive: System Concept and Comparison to Traditional**

**Tools and Emergent Techniques**

**System Creation and Output**

**Business Use Cases**

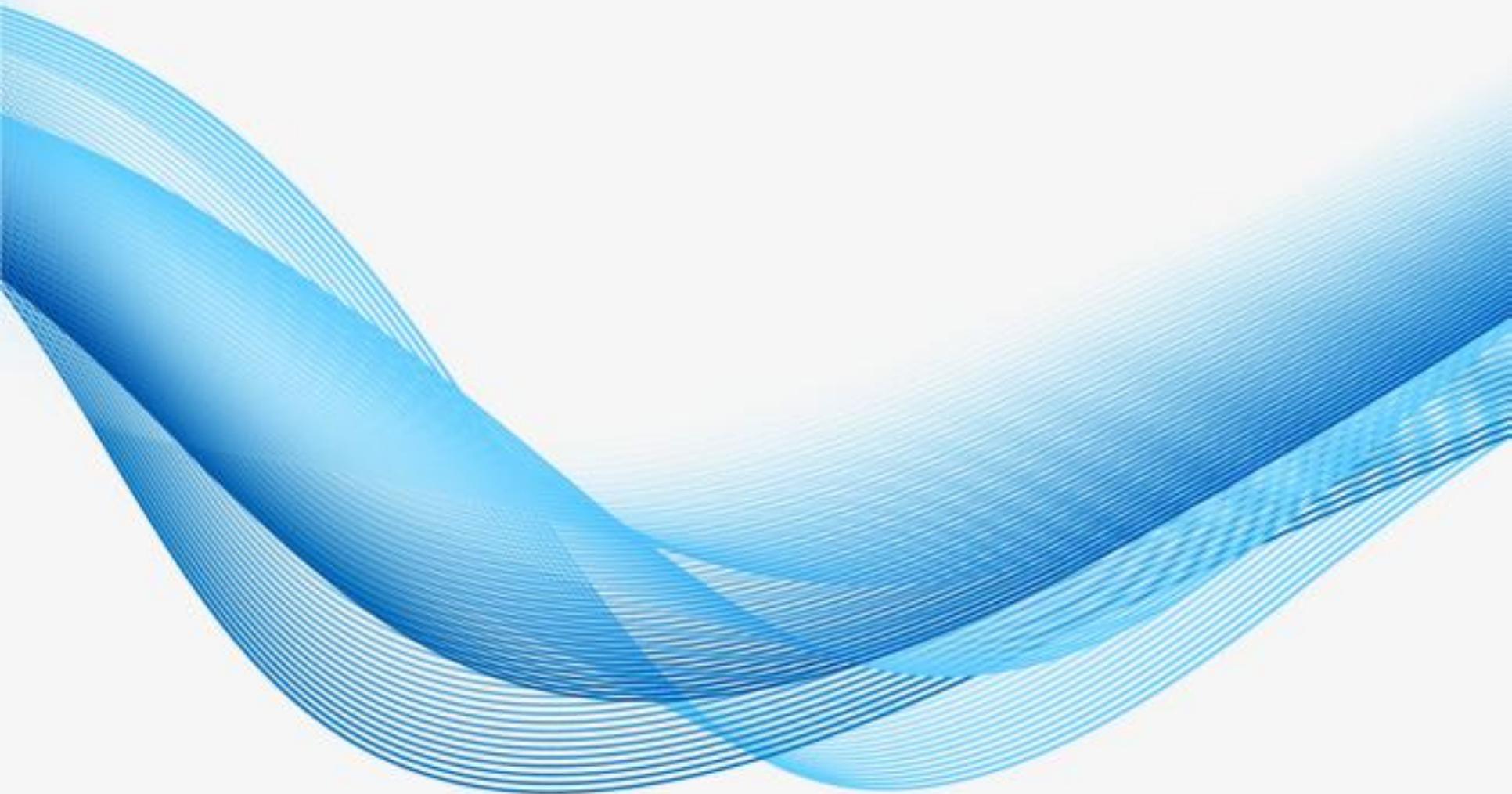
**Conclusion**

# Conclusion

A decorative graphic consisting of several overlapping, wavy blue lines that create a sense of motion and depth, extending across the top of the slide.

- Approach a recommendation problem as a predicative problem
- Leveraging Spark on *AWS* enables business to scale up only when training a model
- Most useful for business challenges with many choices across a significant amount time
- The bigger the data, the smarter the system

# Thank You



**Team 9:** Xiyang Xu, Kitlyn (Hui-Lun) Kuo , Yanlun Ren, Raghuveer Rao Vijjini, Sydney Jiang, Wade Wimer