

Architectures for massive data management

Apache Flink Session Lab

Albert Bifet

albert.bifet@telecom-paristech.fr

université
PARIS-SACLAY



September 22, 2015

Recommendation Engines

- Websites as Amazon help customers to find items that could interest them
- For example: “Customers who bought this item also bought”

Customers Who Bought This Item Also Bought



A screenshot of the Amazon website showing a row of five book recommendations under the heading "Customers Who Bought This Item Also Bought". Each book listing includes a cover image, title, author, star rating, number of reviews, and price with the Prime logo.

Book Title	Author	Rating	Reviews	Price
Predictive Analytics For Dummies	Anasse Bari	★★★★☆	29	\$17.72 ✓Prime
Predictive Analytics: The Power to Predict Who...	Eric Siegel	★★★★☆	229	#1 Best Seller in Econometrics \$16.88 ✓Prime
Quantifying the User Experience: Practical...	Jeff Sauro	★★★★☆	8	\$40.63 ✓Prime
Marketing Analytics: Strategic Models and...	Stephan Sorger	★★★★☆	29	\$50.52 ✓Prime
Data Driven Marketing For Dummies	David Semmelroth	★★★★☆	29	\$20.49 ✓Prime

Recommendation Engines

Customers Who Bought This Item Also Bought



The screenshot displays five book recommendations with their respective covers, titles, authors, ratings, and prices. A left arrow is visible on the far left.

Book Title	Author	Rating	Price
Predictive Analytics For Dummies	Anasse Bari	★★★★☆ 29	\$17.72 ✓Prime
Predictive Analytics: The Power to Predict Who...	Eric Siegel	★★★★☆ 229	#1 Best Seller in Economics Hardcover \$16.88 ✓Prime
Quantifying the User Experience: Practical...	Jeff Sauro	★★★★☆ 8	Paperback \$40.63 ✓Prime
Marketing Analytics: Strategic Models and...	Stephan Sorger	★★★★☆ 29	Paperback \$50.52 ✓Prime
Data Driven Marketing For Dummies	David Semmelroth		Paperback \$20.49 ✓Prime

- Input
 - Set of items in transactions:
book12, book34, cd12, cd42, dvd32
book32, book34, dvd32
- Output using Dataset API
 - For each item what are the items most commonly bought by customers who also bought this item
- Output using DataStream API
 - What are the top k items most commonly bought by customers in general

Apache Flink Session Lab

- Follow the Apache Flink Setup Tutorial
 - https://ci.apache.org/projects/flink/flink-docs-master/quickstart/setup_quickstart.html
- Follow the Apache Flink API Tutorial
 - https://ci.apache.org/projects/flink/flink-docs-master/quickstart/java_api_quickstart.html, OR
 - https://ci.apache.org/projects/flink/flink-docs-master/quickstart/scala_api_quickstart.html
- Write Apache Flink applications in Java or Scala