

Predicting Customer Preferences

John T. Leonard

Krishna Venkatraman

Ross Tomlin

Predicting Customer Preferences | Lessons Learned: Questions Answered

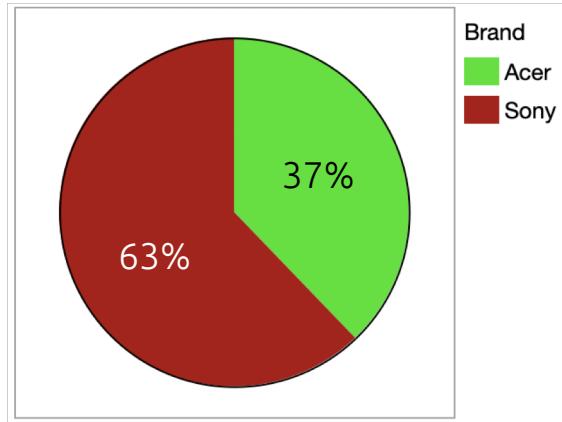
- Which brand do customers prefer:
 - Acer vs. Sony
- Can we predict sales volume for new products?
- Can we recommend products to customers based on product associations (recommender system)
- R vs. RapidMiner?
- Future Usages of Data Analytics

The collage consists of five panels:

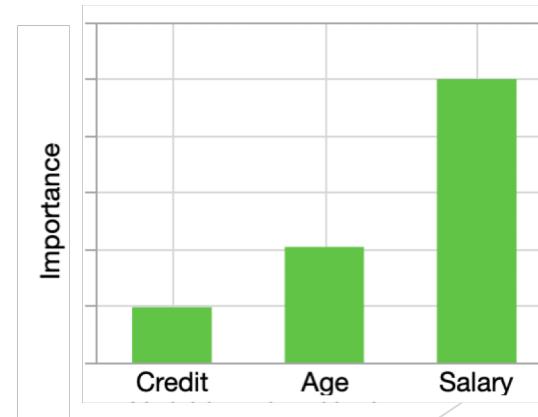
- Panel 1:** A comparison between two computer brands. It features the Acer logo on the left and the Sony logo on the right, separated by the word "VS".
- Panel 2:** A bar chart titled "Total Volume" showing the sales volume of various electronic devices. The categories and approximate volumes are: Laptop (~1), PC (~1), Smartphone (~2), Netbook (~2), Game Console (~4), and Tablet (~4).
- Panel 3:** A diagram illustrating a recommender system. It shows a "Shopping Cart" containing a monitor, a keyboard, and a mouse, followed by a plus sign and a "Frequently Bought With" section featuring a desktop computer.
- Panel 4:** A comparison between R and RapidMiner. It features the R logo on the left and the RapidMiner logo on the right, separated by the word "VS".
- Panel 5:** An illustration of a data mining process. It shows a diver in a wetsuit interacting with a large, glowing, 3D surface that represents a complex dataset or model.

Predicting Customer Preferences | Which Brand do Customers Prefer

- Which brands do customers prefer & why?
- Model predicts customer brand preference with 90% accuracy
- Brand preference primarily depends on 3 variables: Salary, Age, & Credit



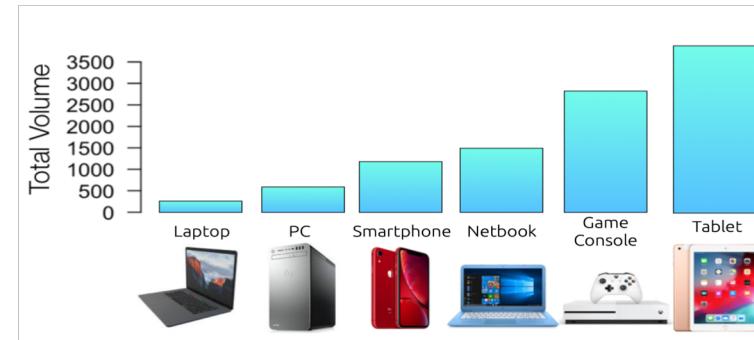
Top 3 Most Important Variables



Predicting Customer Preferences | Predict Sales Volume

- Can we predict sales volume for new products?
- Model predicts sales volume to within +/- 1.8 sales an average for existing products
- Total volume predicted for new products. Top 6 highest volume products shown below

Metric	Sales Volume Avg. Error
Model-1	8.5
Model-2	1.8
Model-3	2



Discover Product Associations

Market Basket Analysis

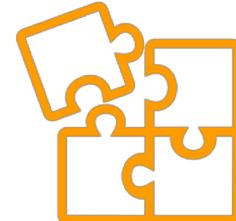
- Discovering relationships
- Association of Customer transactions & purchased items
- Uncover insights on transactional datasets

Measurements

- Support & Confidence
- Strong & Insightful Rules
- Irrelevant & Unclear Rules

Problem Definition

Blackwell Electronics intent to acquire Electronidex



Discovered Insight

Blackwell to acquire Electronidex to amplify cross-selling & net profitability

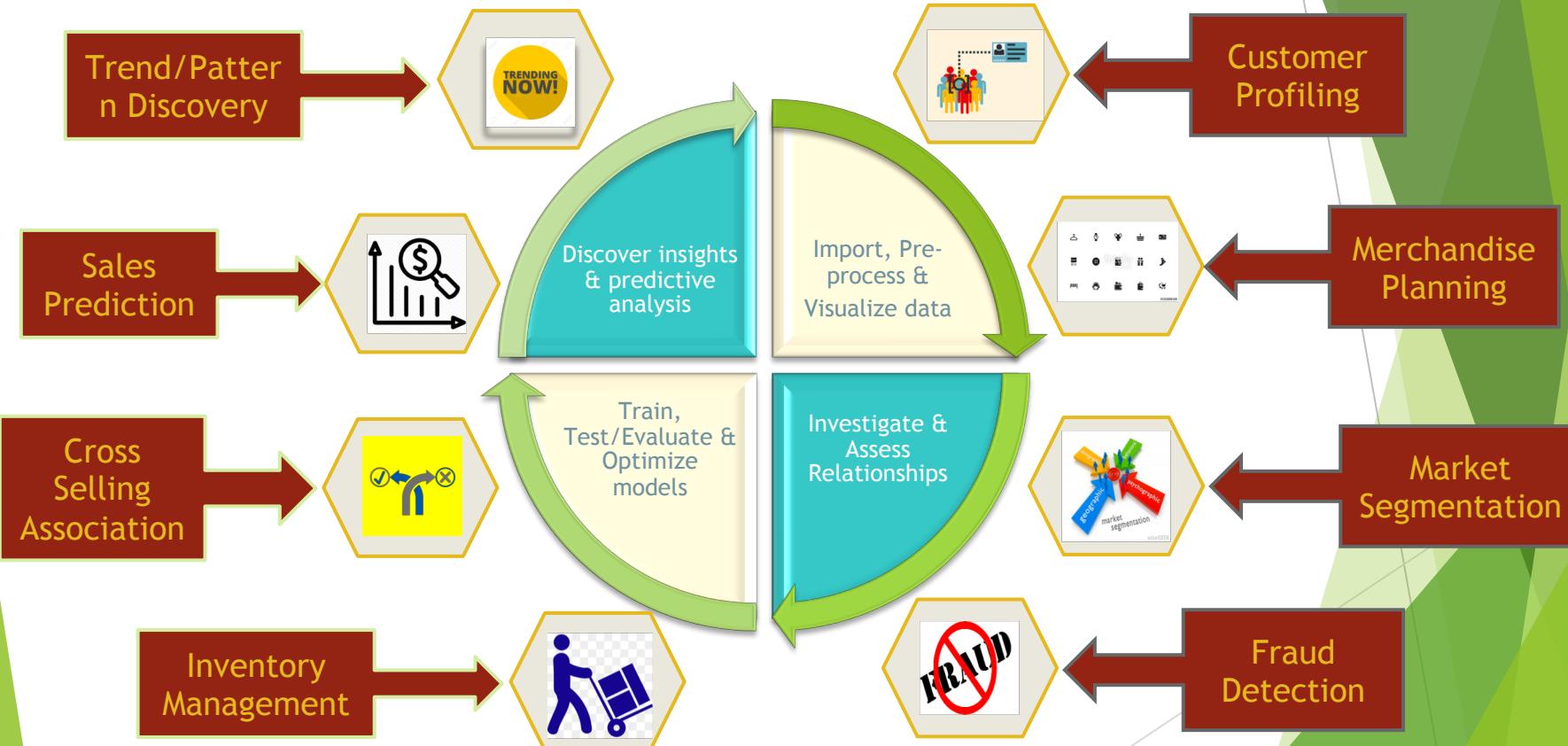
Metric	
# of Rules	197
Mean Support	0.93
Mean Confidence	0.82
Mean Lift	0.10

R vs Rapid Miner Comparative Analysis



Description	Software programming language & software environment for statistical computing and graphics.	Software platform IDE provides an integrated environment for machine learning, data mining, text mining, predictive analytics & business analytics
Price	Open Source	Free Community Edition & Commercial Enterprise Edition
Language	R interpreted Language	GUI based IDE developed using JAVA
input/output files format	text file (ASCII,.dat), Excel spreadsheet (.csv,.delim,.DIF), Pdf, html, images,minitab, s-plus	text file (ASCII,.dat), Excel spreadsheet (.csv,.delim,.DIF), Pdf, html, images, audio, SAP
Visualization Plots	Bar chart, Pie Chart, Line, Histogram, Scatter, Conditioning Plot, Kernel Density, Quantile Quantile, Plotly, Google Visualization API	Bar chart, Pie Chart, Line, Histogram, Scatter, Bubble, Deviation, Survey plots, Density, Andrews Curves, Quartile
Performance	Best accuracy levels with missing values in datasets	Best to handle continues data type & multi class data
Structure & Development	Built-in and user-created packages & functions and in-memory data objects. Extensible using R code and Generics	Built using normal operators containing one or more sub processes & super operators relying on execution of other user defined operators. Extensible using built-in scripting operator & extensions providing new operators and new data objects

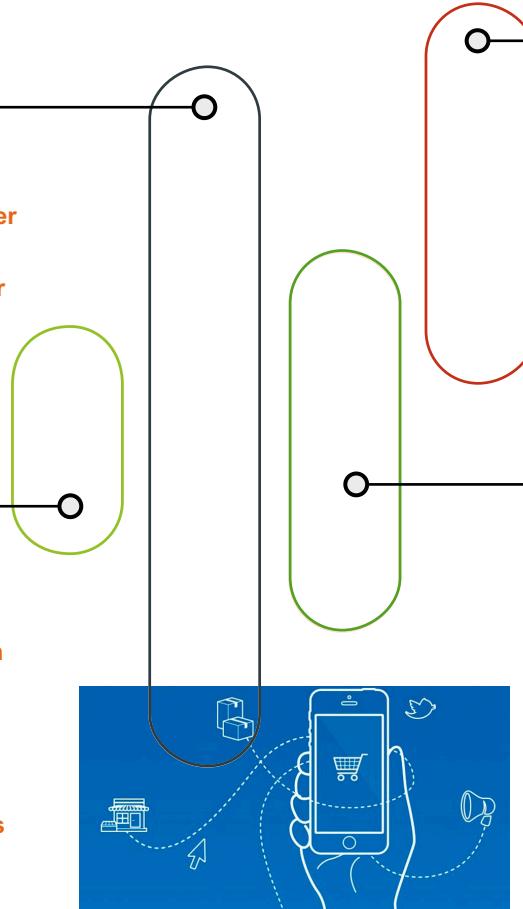
Applications of Data Mining



Blackwell's Flight Journey to future selling

Customer centric User Experience

- ✓ Customers want differentiated experiences rather than commodity products
- ✓ Blackwell to focus around positive outcomes for their customers rather than simply moving merchandise
- ✓ Improvement in Customer experience with 72 % Confidence



Analytics Driven Sales Prediction Strategy

- ✓ Learn about potential customers & figure out exactly what they want
- ✓ Learn how to guide them through evermore complex sales cycles
- ✓ manage inventory, supply chain, forecasting demands & pricing strategies

Self-Service Capability

- ✓ Mobile App/ Website immaculately designed to give potential customers a raft of information on products, reviews and articles about Blackwell and easy paths to purchase, all wrapped in a cleanly designed, package
- ✓ It's as if the company's website & Mobile Apps are its sixth product, created to make customers happy

Delivery at Speed

- ✓ Ease of online payments
- ✓ Increased focus on Micro Segments
- ✓ Offer quick actions for mobile & website searches on—I want to go, I want to know, I want to buy, etc.

Questions?

Appendix

Predicting Customer Preferences | Lessons Learned: *Introduction & Overview*

- Questions Answered
- Methods
- Results
- Future Opportunities



Predicting Customer Preferences | Lessons Learned: Questions Answered

- R vs. RapidMiner?
- Which brand do customers prefer:
 - Acer vs. Sony
- Can we predict sales volume for new products?
- Can we recommend products to customers based on product associations (recommender system)



Predicting Customer Preferences | Lessons Learned: *Methods*

- 3 classes of machine learning models/methods used:
 1. Classification: Which brand do customers prefer
 - C5 decision tree, Random Forest
 2. Regression: Can we predict sales for new products
 - support vector machine (SVM), Random forest, Gradient boosted tree
 3. Apriori

Predicting Customer Preferences | Lessons Learned: Methods

- 3 classes of machine learning models/methods used:
 1. Classification: Which brand do customers prefer
 - C5 decision tree, Random Forest
 2. Regression: Can we predict sales for new products
 - support vector machine (SVM), Random forest, Gradient boosted tree
 3. Apriori

Brand Preference Classifier Models

Metric	Accuracy	Kappa
Decision Tree	0.92	0.84
Random Forest	0.89	0.87

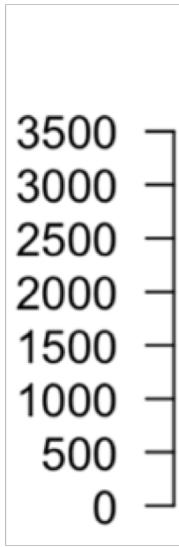
New Product Sales Volume Prediction Models
(Regression)

Metric	RMSE $\log_{10}(\text{Volume})$	R ²
SVM	0.93	0.10
Random Forest	0.28	0.89
Gradient Boosted Tree	0.30	0.88

Apriori

Metric	
# of Rules	197
Mean Support	0.93
Mean Confidence	0.82
Mean Lift	0.10

Total Volume



Laptop



PC



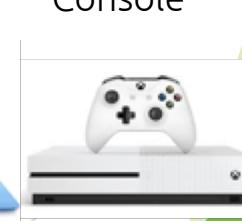
Smartphone



Netbook

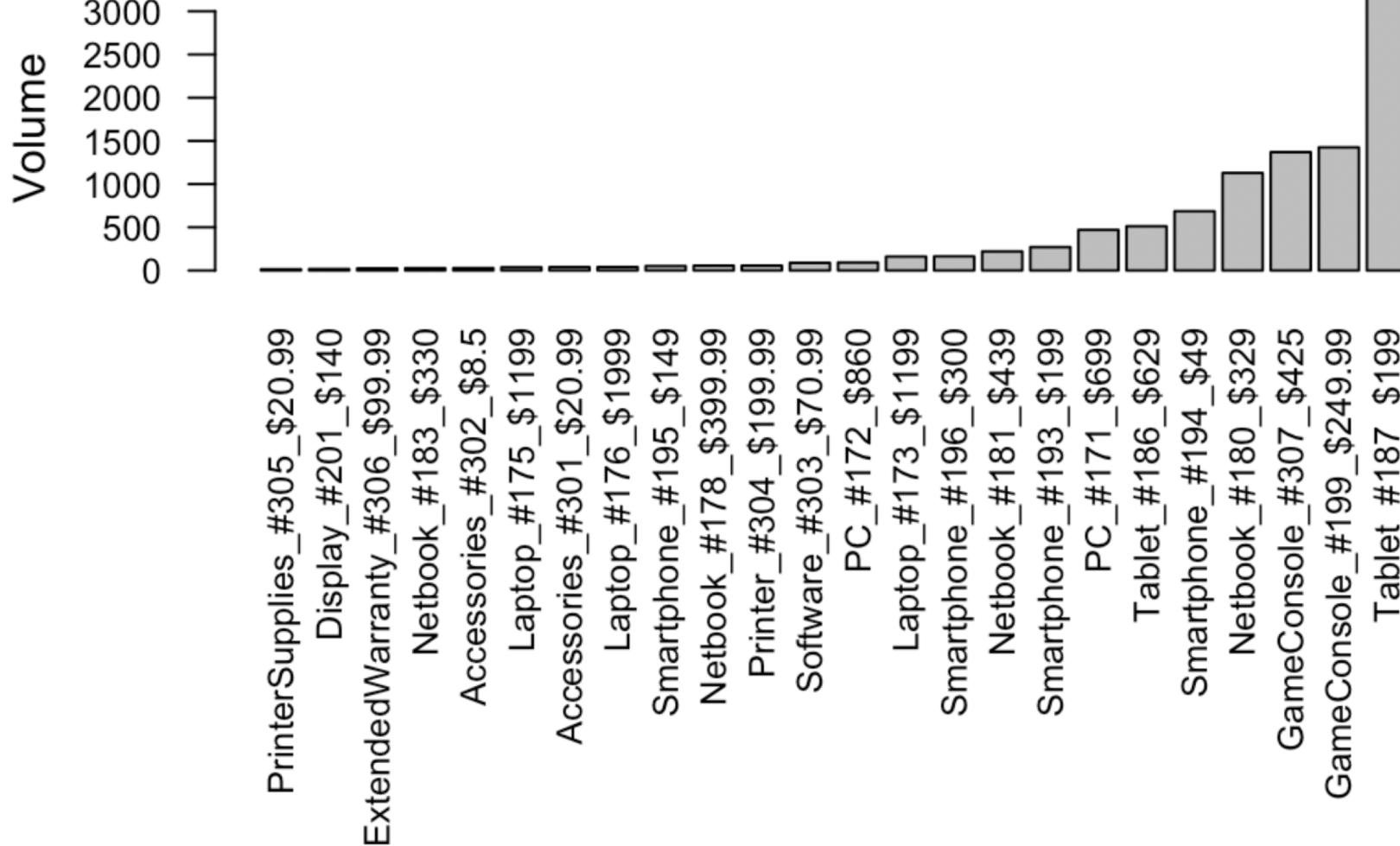


Game Console



Tablet





Rules		Support	Confidence	Lift	count
{Dell Desktop,iMac,Lenovo Desktop Computer,Mackie CR Speakers} => {ViewSonic Monitor}		1.12E-03	1.00	9.1	11.00
{Belkin Mouse Pad,Microsoft Office Home and Student 2016,Rii LED Gaming Keyboard & Mouse Combo} => {Dell Desktop}		1.02E-03	0.91	6.8	10.00
{ASUS Monitor,Intel Desktop,ViewSonic Monitor} => {Lenovo Desktop Computer}		1.12E-03	1.00	6.8	11.00
{Dell KM117 Wireless Keyboard & Mouse,iPhone Charger Cable} => {Apple MacBook Air}		2.03E-03	0.95	6.1	20.00
{Acer Aspire,Koss Home Headphones,ViewSonic Monitor} => {HP Laptop}		1.22E-03	1.00	5.2	12.00
{Brother Printer,Halter Acrylic Monitor Stand} => {iMac}		1.12E-03	1.00	3.9	11.00



Train-Test Set	Train+Test		Train		Test	
Metric	Accuracy	Kappa	Accuracy	Kappa	Accuracy	Kappa
Decision Tree	0.93	0.85	0.94	0.85	0.92	0.84
Random Forest	0.97	0.94	0.99	0.99	0.89	0.87

Metric	RMSE $\log_{10}(\text{Volume})$		R^2	
Train-Test Set	Train	Test	Train	Test
SVM	0.32	0.93	0.82	0.10
Random Forest	0.09	0.28	0.98	0.89
Gradient Boosted Tree	0.01	0.30	0.99	0.88

Metric	# Rules	Mean Support	Mean Confidence	Mean Lift
Apriori	197	0.93	0.82	0.10

Metric	Accuracy	Kappa	Quality
Model 1	0.92	0.84	0.88
Model 2	0.89	0.87	

Variable Importance for Random Forest: from the R package: “For each tree, the prediction accuracy on the out-of-bag portion of the data is recorded. Then the same is done after permuting each predictor variable. The difference between the two accuracies are then averaged over all trees, and normalized by the standard error. For regression, the MSE is computed on the out-of-bag data for each tree, and then the same computed after permuting a variable. The differences are averaged and normalized by the standard error. If the standard error is equal to 0 for a variable, the division is not done.”

Top 3 Most Important Variables

Random Forest Variable Importance:	
	Overall <dbl>
salary	100.0000000
age	40.7899450
credit	19.5454770
level3	0.8989291
level2	0.8151075
level1	0.8132614
level4	0.7608397
zipcode4	0.4374877
zipcode5	0.4051055
zipcode7	0.3951130

