Big Data in Banking: Hype or Future?

prof. David Martens
david.martens@uantwerp.be
www.applieddatamining.com
Agenda

- What’s new about “Big” Data?
  - Marketing applications in banking
    - Response modeling using payment data
    - Customer acquisition using browsing data
  - Risk management applications in banking
    - Retail default prediction using Facebook data
    - SME Default prediction using board of director data
- Challenges for the future in banking
SAY BIG DATA
ONE MORE TIME
What is Big Data?
What is Big Data?

**Big Data**: data that is so large that traditional data processing systems are unable to deal with it

- Storage and analysis
What is Big Data?

- Hadoop
  - Open-source framework for data-intensive distributed processing on commodity hardware
  - Derived from MapReduce and Google File System (GFS) papers

- Big Data > Hadoop
What is Big Data?

- **Google**
  - 24,000 TB per day (2009)

- **Pinterest**
  - 20 TB per day

- **Twitter**
  - 12 GB per day or 800 tweets per second (2010)

- **PrediCube, spinoff UA**
  - 30 GB per day

- **Bank payment data**
  - 5-10 GB for all payment data of one year

*Is Banking data really Big?*
What is Data Mining?

- **Data mining**: automatic extraction of knowledge from data
- Setting the scene with credit scoring example

---

**Data**

**Data mining technique**

**Pattern**

**Predictions**

<table>
<thead>
<tr>
<th>Client</th>
<th>Income</th>
<th>Sex</th>
<th>Amount</th>
<th>Default</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>1,600</td>
<td>M</td>
<td>175,000</td>
<td>N</td>
</tr>
<tr>
<td>B</td>
<td>2,600</td>
<td>F</td>
<td>350,000</td>
<td>Y</td>
</tr>
<tr>
<td>C</td>
<td>3,280</td>
<td>M</td>
<td>50,000</td>
<td>N</td>
</tr>
<tr>
<td>D</td>
<td>950</td>
<td>M</td>
<td>120,000</td>
<td>Y</td>
</tr>
<tr>
<td>E</td>
<td>10,500</td>
<td>M</td>
<td>1,000,000</td>
<td>N</td>
</tr>
<tr>
<td>F</td>
<td>5,700</td>
<td>F</td>
<td>240,000</td>
<td>N</td>
</tr>
<tr>
<td>G</td>
<td>2,400</td>
<td>F</td>
<td>250,000</td>
<td>N</td>
</tr>
</tbody>
</table>

---

**Classification Model**

If income < 10,000 and Amount Loan > 100,000 and ...
then default = yes
## What is Big Data?

### Transaction II

<table>
<thead>
<tr>
<th>Transaction ID</th>
<th>Items</th>
</tr>
</thead>
<tbody>
<tr>
<td>0001</td>
<td>Bread, Milk, Apple</td>
</tr>
<tr>
<td>0002</td>
<td>Bread, Milk, Eggs, Pen</td>
</tr>
<tr>
<td>0003</td>
<td>Cold Drink, Chocolate, Milk</td>
</tr>
<tr>
<td>0004</td>
<td>Bread, Orange</td>
</tr>
<tr>
<td>0005</td>
<td>Fish, Vegetables</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td>0052</td>
<td>Paper, Pencil</td>
</tr>
<tr>
<td>0053</td>
<td>Meat, Oil, Milk</td>
</tr>
</tbody>
</table>

### Classification Model

```plaintext
if income < 10,000 and Amount Loan > 100,000 and ... then default = yes
```

### Frequently Bought Together

- **Total List Price:** $225.99
- **Price For All Three:** $170.06

**This Item:** Introduction to Data Mining by Pang-Ning Tan

**Data Mining:** Practical Machine Learning Tools and Techniques, Second Edition (Morgan Kaufmann Series in Data Management Systems) by Ian H. Witten

**Data Mining:** Concepts and Techniques, Second Edition (The Morgan Kaufmann Series in Data Management Systems) by Micheline Kamber Jiawei Han
What is Big Data?

Banking

- The story of Signet Bank
  - 1990
  - Fairbanks and Morris
  - Model profitability, not just default
  - No interest from big banks
  - Signet Bank invested in data assets
  - Huge success, credit operations spinoff
    ➔ Now Capital One

**Defining and leveraging data assets**
Agenda

- What’s new about “Big” Data?

  - **Marketing** applications in banking
    - Response modeling using payment data
    - Customer acquisition using browsing data

  - **Risk management** applications in banking
    - Retail default prediction using Facebook data
    - SME Default prediction using board of director data

- **Challenges** for the future in banking
Case Study 1: Mining Payment Data

- From payment data to pseudo-social network
  - 21 million transactions
  - Response modeling
  - Significant improvement over traditional modeling

Payment receivers

customers

John
Alex
An
Pete
Jeff

Little Bookstore
DeliC
Amazon
SportCenterX
EnergyInc

[David Martens, Foster Provost, Pseudo-social network targeting from consumer, transaction data, New York University - Stern School of Business - Working paper CeDER-11-05
Patent application PCT/US2011/028175]
Case Study 1: Mining Payment Data

- From payment data to pseudo-social network
  - 21 million transactions
  - Response modeling
  - Anonymized!
Case Study 1: Mining Payment Data

- Application: response modeling
  - Target variable on two products
  - Pension fund and Long term deposit account
- ‘Socio-demographic’ data
  - 289 variables
  - Socio-demographic, product possession, product use, customer behavior
- When sending offer to top 1%: **3 times more conversions**!

*Variety of (already available) data improves performance*
Case Study 1: Mining Payment Data
Is Bigger Data Better?

SD: Using bank’s structured data

Size of data set used for training (% of 1.2 million consumers in total)

Bigger not better!

Case Study 1: Mining Payment Data
Is Bigger Data Better?

Bigger is better!

PSN: prediction based on data on fine-grained behavior
SD: traditional predictive modeling based on socio-demographic data
PSN + SD: ensemble model combining both

Volume of (already available) data improves performance
Case Study 2: Customer Acquisition

- Predict product interest based on web browsing data
  - Show ad only to those that are predicted to be interested
  - Spinoff of UA

- Data: 1 billion records of persons visiting webpages

- Results: +20-300% conversions

Work done with Dieter Devlaminck (PrediCube)  
www.predicube.com
Agenda

- What’s new about “Big” Data?

- **Marketing** applications in banking
  - Response modeling using payment data
  - Customer acquisition using browsing data

- **Risk management** applications in banking
  - Retail default prediction using Facebook data
  - SME Default prediction using board of director data

- **Challenges** for the future in banking
Case Study 3: Default prediction with Facebook data

- Default prediction with Facebook data for micro-finance
  - In collaboration with NY-based Lenddo
  - Philippines, 1000 – 5000 $ loans
  - Facebook data (opt-in)

Work done with Sofie De Cnudde, Ellen Tobback, Julie Moeyersoms, Marija Stankova (Universiteit Antwerpen) and Vinayak Javaly (Lenddo)
Case Study 3: Default prediction with Facebook data
Friends network

Cluster of befriended non-defaulters

Cluster of befriended defaulters
Case Study 3: Default prediction with Facebook data

Liking data
Case Study 3: Default prediction with Facebook data

Facebook data is very predictive for default prediction.

Behavioral data is more valuable than social network data.
Case Study 4: SME default prediction

- SME default prediction
- Data (Belfirst)
  - 400,000+ SMEs
  - Financial + Board members and managers

Network of companies

Bottom node projection

Work done with Ellen Tobback, Julie Moeyersoms, Marija Stankova (Universiteit Antwerpen)
Case Study 4: SME default prediction

Predictions based on profile of connected companies

Example network of connected defaulting SMEs, due to two directors
Case Study 4: SME default prediction

Big Data?

Just 1% of the data
Case Study 4: SME default prediction

Sample network of connected defaulting SMEs, with three clusters due to three persons
Case Study 4: SME default prediction

- Predictive power for default prediction
  - If we consider the 1% most riskiest SMEs, we can find 58% more defaulters
Agenda

- What’s new about “Big” Data?

- **Marketing** applications in banking
  - Response modeling using payment data
  - Customer acquisition using browsing data

- **Risk management** applications in banking
  - Retail default prediction using Facebook data
  - SME Default prediction using board of director data

- **Challenges** for the future in banking
Privacy!

- “Hey, you’re having a baby!” Target
Privacy vs data as an asset - a spectrum

- Using and selling all data
- Selling payment data
- Selling socio-demo data
- Selling anonymised data
- Using Facebook data
- Tracking users online
  - Using anonymized payment data for marketing
  - Using anonymized payment data for risk management
- Using socio-demo and credit data for risk management
- Nothing
Competition from Google & Co?

- Mainly in payment Data ➔ *Fees and data asset*

- Leveraging their existing data asset

> Though it’s unclear whether Facebook will ever use this patent for that application
Competition from FinTech?

- Niche products
- Loyal Belgians: ability of banks to retain their customers
- FinTech startups: collaborate or acquire

**Percentage of respondents who purchased at a bank other than their primary bank, 2014**

- Developing country average: 47%
- Developed country average: 28%

Source: Bain/Research Now NPS surveys, 2014
References

- **Data Science for Business**
  - By Foster Provost and Tom Fawcett
  - O’Reilly

- **Predictive Analytics: Techniques and Applications in Credit Risk Modelling**
  - By Tony van Gestel, Bart Baesens and David Martens
  - Oxford University Press
Conclusion

- Big Data in Banking: Hype or Future?
  - Both!

- Big Data
  - Banks already quite advanced in data analyses
  - Don’t be fooled by the hype

- What to do?
  - Define and leverage data assets!
  - Collaborate with emerging FinTech startups
  - Main challenge: attracting data scientists
Q&A

Prof. dr. ir. David Martens
Applied Data Mining
Faculteit Toegepaste Economische Wetenschappen
Universiteit Antwerpen

E: david.martens@uantwerpen.be
T: @ApplDataMining
W: www.applieddatamining.com