Customer Care Services System (CCSS)

The data-mining approach for customer care services in grocery shopping

Presenter: Liliana Quyen Tang

COMP 602 – Fall 2018

Presentation Outline

- What is data mining?
- What is Customer Care Services System (CCSS)?
- Why should we use data mining for CCSS?
- How do we use data mining in CCSS?
- How data mining in CCSS can benefit from R vs Python vs SQL?

What is data mining?

- “Solve problems by analyzing data already present in databases” (Witten & Frank, 2005)
- Is all about seeking patterns automatically
- Today: we try to make sense of customer behaviours in retail grocery shopping

What is Customer Care Services System (CCSS)?

- A desktop application used in grocery stores
- Takes data from Point-Of-Sale System, then applies data mining techniques to seek out useful hidden information
- Is similar to Customer Relationship Management (CRM) platform
- Includes:
  - Rule-mining based systems (Rategari & Sap, 2008)
  - Event-prediction systems (Rategari & Sap, 2008)
  - Web personalization systems (Rategari & Sap, 2008)

Why should we use data mining for CCSS?

- Improve
  - Customer satisfaction
  - Prediction ability
    - Profit
    - Sustainability
- Reduce
  - Waste (time, financial, and human resources)
  - Damage to the product due to natural impact, eg. Snow, rain, etc

How do we use data mining in CCSS?

- Common techniques in data mining (Witten & Frank, 2005):
  - Association Rules:
    - Application: Market basket analysis for supermarket checkout data
    - Example: Customer who buy beer also buy chips
  - Classification Rules
    - Application: Decision trees – “buy” or “not buy”
    - Example: People who are older than 40 years old usually don’t prefer contact lenses over glasses with frame
  - Clustering - Group data that is likely to fall naturally together
    - Application: used to predict outliers such as shoplifting or customers who no longer go shopping in the store

How data mining in CCSS can benefit from R vs Python vs SQL?

- Willems 2015:
  - "What problems do you want to solve?"
  - "What are the net costs for learning a language?"
  - "What are the commonly used tool(s) in your field?"
  - "What are the other available tools in your field and how do these relate to the commonly used tool(s)?"

- SQL (Wang 2016)
  - Is necessary for querying and extracting data -> 1st step to get the data into usable format (eg. JOIN Statements) & for complex operations
- Python (Wang 2016)
  - Is crucial for manipulating or transforming data (eg. Statistical analysis, regressions, trend lines, and time series data) & does offer some libraries for visualizing purposes
- R (Willems 2015)
  - Is easy to learn for people without programming background
  - Is perfect choice for visualization, but quite slow

Benn Stancil, Chief Analytics Officer at Mode

- SQL

```
 WITH details AS (
 SELECT series,
        value,
        ROW_NUMBER() OVER (PARTITION BY series ORDER BY value) AS row_number,
        SUM(1) OVER (PARTITION BY series) AS total
   FROM dataset
 ),

 quartiles AS (
 SELECT series,
        value,
        AVG(CASE WHEN row_number >= (FLOOR(total/2.0)/2.0)
                  AND row_number <= (FLOOR(total/2.0)/2.0) + 1
                 THEN value/1.0 ELSE NULL END
           ) OVER (PARTITION BY series) AS q1,
        AVG(CASE WHEN row_number >= (total/2.0)
                  AND row_number <= (total/2.0) + 1
                 THEN value/1.0 ELSE NULL END
           ) OVER (PARTITION BY series) AS median,
        AVG(CASE WHEN row_number >= (CEIL(total/2.0) + (FLOOR(total/2.0)/2.0))
                  AND row_number <= (CEIL(total/2.0) + (FLOOR(total/2.0)/2.0) + 1)
                 THEN value/1.0 ELSE NULL END
           ) OVER (PARTITION BY series) AS q3
   FROM details
 )
SELECT series,
       MIN(CASE WHEN value >= q1 - ((q3-q1) * 1.5) THEN value ELSE NULL END) AS minimum,
       AVG(q1) AS q1,
       AVG(median) AS median,
       AVG(q3) AS q3,
       MAX(CASE WHEN value <= q3 + ((q3-q1) * 1.5) THEN value ELSE NULL END) AS maximum
 FROM quartiles
 GROUP BY 1
```

- Python

Dataset.describe()

References

- Griva, A., Bardaki, C., Pramatari, K. & Papakiriakopoulos, D. (2018). Retail Business Analytics: Customer Visit Segmentation Using Market Basket Data. Expert Systems with Applications, 100 (2018), 1-16
- Min, H. (2006). Developing the Profiles of Supermarket Customers through Data Mining. The Service Industries Journal, 26)7), 747-763
- Pan, S., Giannikas, V., Han, Y., Grover-Silva, E. & Qiao, B. (2017). Using customer-related data to enhance e-grocery home delivery. Industrial Management & Data Systems, 117(9), 1917 – 1933
- Rategari, H. & Sap, M. (2008). Data mining and e-commerce: methods, applications, and challenges. Jurnal Teknologi Maklumat, 20(2), 116 – 128
- Wang, L, (2016). What SQL Analysts Need to Know About Python. Segment. https://segment.com/blog/sql-vs-python/. Accessed 03.12.2018
- Willems, K. (2015). Choosing R or Python for Data Analysis? An Infographic. DataCamp. https://www.datacamp.com/community/tutorials/r-or-python-for-data-analysis. Accessed 03.12.2018
- Witten, I. & Frank, E. (2005). Data mining: practical machine learning tools and techniques. 2nd ed. San Francisco, CA: Elsevler
- Yang, Y., Liu, H. & Cai, Y. (2013). Discovery of Online Shopping Patterns Across Websites. INFORMS Journal on Computing, 25(1), 161 – 176

Questions?

Thank you for watching!