BICOL UNIVERSITY COLLEGE OF SCIENCE

CS Elective – Data Mining

Case Study #1

CANONIZADO, MICHAEL XAVIER, E.

Part II. Post Case Study: *"SmartMart's Customer Insight Initiative"*

Case Background

SmartMart is a medium-sized retail chain operating both physical stores and an online shopping platform. Over the past five years, the company has accumulated a **large database of customer transactions**, loyalty card records, mobile app usage logs, and customer feedback.

Management believes that hidden patterns in this data could help:

- Improve product placement
- Predict customer demand
- Personalize promotions
- Reduce inventory waste

To achieve this, SmartMart plans to adopt **data mining techniques** and has formed a small analytics team composed of IT staff, business analysts, and consultants.

However, concerns have been raised regarding:

- The privacy of customer information
- The accuracy and reliability of extracted patterns
- Whether the company fully understands the difference between data mining, statistics, and AI
- The tools and skills needed to responsibly implement the initiative

You are part of an advisory group asked to evaluate the plan.

1. Understanding Data Mining

a. Based on the case, explain what data mining means in your own words.

b. Why is data mining more than simply storing or querying data?

- Data Mining is extracting patterns from large datasets. Storing and querying data is just the first step, data mining is all about discovering meaning and turning data into information.

2. Data Mining Tasks. Identify which data mining tasks are applicable in SmartMart's case.
For each, briefly explain how it could be used.

- Classification

  Predict whether people will actually buy the items that we will put on sale based on historical data
- Clustering

  This can be used to classify the customers into groups based on their purchasing patterns. Such as people who like to buy gadgets, food, people who only buy on-sale items, etc.
- Association Rule Mining

  This can help us position the items within the store appropriately. Such as putting popular items eye-level, or putting related items far apart so the customers would have to spend more time at the store.
- Prediction / Forecasting

  We can predict the sales for a season, such as christmas and black-friday. Therefore we can adjust our inventory appropriately.
- Anomaly Detection

  This can help us spot fraudulent, system errors, or"fishy" cases. Such as high bills (e.g 50k php) when the average is only 3k, suspicious payment methods, or uncommon item combinations.

3. Relationship to Other Fields

- Explain how SmartMart's initiative relates to the following fields:
-

Statistics: Main math foundation in SmartMart's initiative. This is the backbone of data-mining that turns raw data into something we can understand and trust. We can get core information such as averages, median, distributions, etc. which helps us understand the data. This can be then used for hypothesis creation, analytics, and more.

Optimization: Finding the best solution among many solutions. This helps SmartMart find the most efficient algorithms and techniques in things like: model training, feature selection, clustering, etc.

Machine Learning: Automating pattern discovery and predictions from SmartMart's large datasets. ML provides the algorithms that can classify data, predict values, anomaly detection, etc.

Artificial Intelligence: Data mining results can be used in AI to utilize the information and patterns gathered. It can do things like enhance their digital store where they can perform things such as real time recommendations on the digital store.

4. Tools and Technologies

a. What types of data mining tools would be appropriate for SmartMart (e.g., spreadsheet-based, open-source, enterprise tools)?

- Since SmartMart has large and diverse datasets, they need tools that can handle this volume and complexity. Tools such as SQL, Pandas, Tableau, Weka, etc.

b. Why might Python-based tools be preferred over traditional spreadsheet tools as data size grows?

- Python-based tools are perfect for large growing datasets, as it can process millions of records, automated scripts, and is reliable.

5. Data Considerations

a. Identify at least three types of data used in the case.

1) Structured Data from their system databases.
2) Transactional Data from their store purchases.
3) Semi-Structured data also from their systems.

b. What data quality issues might affect the accuracy of results?

- Incomplete data, inconsistent data, duplicate entries, outdated data, etc. All of these needs to be processed before data-mining.

c. Why is data preprocessing critical before applying data mining techniques?

- It is critical to process the data before performing any data-mining techniques to get the best and most accurate results. It ensures that the results are consistent and has little noise.

6. Legal, Privacy, and Security Issues

a. What privacy risks arise from analyzing customer behavior data?

- Analyzing their customer data could pose some privacy risk such as: Unauthorized access, data breaches if their system is not secure, and excessive collection of customer data.

b. What ethical boundaries should SmartMart observe when using customer data?

- To maintain customer trust, SmartMart needs to operate on data responsibly. They need to be transparent and get the consent of customers.

c. How can data mining cross the line into unethical or illegal practice?

- Data mining can be unethical and illegal when it does things like: Using data without consent, no data-protection and security, and overall lacking legal compliance.

7. Decision-Making Reflection

a. Should SmartMart proceed with its data mining initiative? Why or why not?

- Yes SmartMart should process with their data mining initiative. It will greatly boost their business value, and all their goals such as improving product placement, demand forecasting, etc. are all possible with data mining. However, they should still proceed with caution, ensuring that they have legal compliance and safeguarding their customers' data.

b. What safeguards should be implemented before deployment?

- Have clear policies on data ownership and access.
- Have privacy and legal compliance.
- Security protection on customer data
- And properly trained staff who will operate on the data.