BICOL UNIVERSITY COLLEGE OF SCIENCE

CS Elective – Data Mining

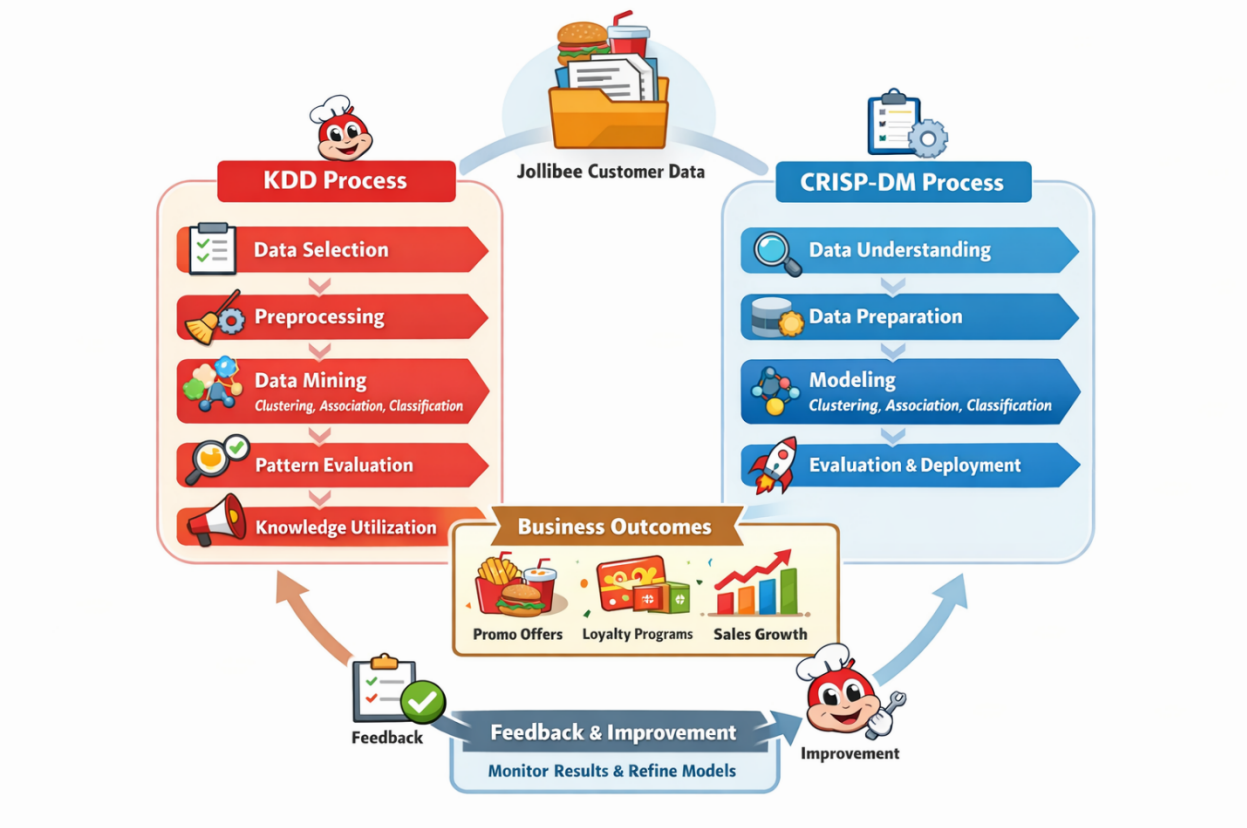
Case Study #1

LN, FN, MI.

PART 1: Post-Exercise Applying KDD/CRISP-DM to Jollibee

Instructions

1. Choose KDD or CRISP-DM as your framework.
2. Identify the data features you will use from a Jollibee dataset.
3. Suggest at least one descriptive task (e.g., clustering or summarization) and one predictive task (e.g., classification or regression).
4. Outline how you would present the results to Jollibee management.
5. Propose how feedback will be collected to improve future campaigns.



Jollibee wants to improve customer loyalty and upselling by understanding purchasing patterns across its food outlets.

Option A: Using KDD Framework

1. **Data Selection**

Select relevant data:

- Customer orders (items, quantity, total price)
- Time and day of order
- Store location
- Loyalty card usage

2. **Data Preprocessing**

- Remove duplicates or cancelled orders
- Handle missing data (e.g., customer ID not recorded)
- Standardize product names

3. **Data Transformation**

- Aggregate total spending per customer per month
- Encode menu categories (e.g., burgers, drinks, desserts)
- Create "time of visit" feature (breakfast, lunch, dinner)

4. **Data Mining**

- **Clustering:** Group customers by purchase frequency and spending
- **Association Rules:** Discover items frequently bought together
- **Classification/Prediction:** Predict if a customer will buy a new product

5. **Pattern Evaluation**

- Evaluate rules by support and confidence
- Check clusters for business relevance

6. **Knowledge Presentation**

- Present dashboards: frequent combos, customer segments, sales trends

7. **Knowledge Utilization**

- Suggest promotions to high-value segments
- Design meal combos for upselling

8. **Feedback**

- Monitor sales and promo responses
- Refine patterns for future campaigns

Option B: Using CRISP-DM Framework

1. **Business Understanding**

- Goal: Increase sales and customer loyalty
- Success metric: Higher repeat visits, increased average order value

2. **Data Understanding**

- Explore historical sales, customer visits, popular menu items
- Identify missing or inconsistent data

3. **Data Preparation**

- Clean, transform, and aggregate features
- Example: total monthly spending, favorite menu category, visit frequency

4. **Modeling**

- Apply clustering to segment customers
- Association rules to identify popular combos
- Predictive models to forecast new product adoption

5. Evaluation

- Check model accuracy
- Ensure patterns make business sense
- Identify actionable insights

6. Deployment

- Use insights for marketing campaigns
- Personalized combo offers and promotions
- Loyalty program targeting

Example Dataset

CustomerID	VisitsPerMonth	AvgOrderValue (₱)	FavoriteCategory	BoughtCombo?
001	5	250	Burger	Yes
002	2	150	Chicken	No
003	7	300	Burger	Yes
004	1	100	Desserts	No
005	4	200	Burger	Yes

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?

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
- Clustering
- Association Rule Mining
- Prediction / Forecasting
- Anomaly Detection

3. Relationship to Other Fields

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

Field	Role in the Case
Statistics	
Optimization	
Machine Learning	
Artificial Intelligence	

4. Tools and Technologies

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

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

5. Data Considerations

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

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

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

6. Legal, Privacy, and Security Issues

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

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

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

7. Decision-Making Reflection

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

b. What safeguards should be implemented before deployment?