BICOL UNIVERSITY COLLEGE OF SCIENCE

CS Elective – Data Mining

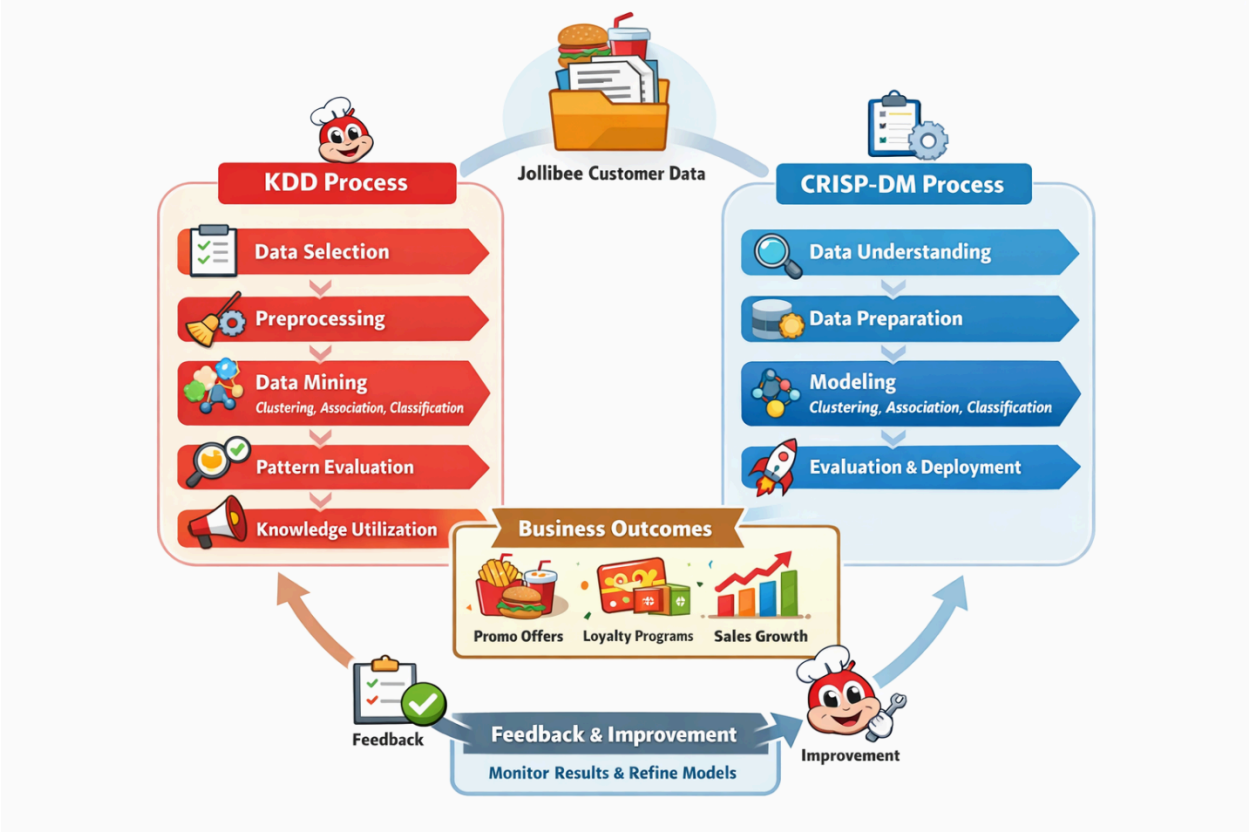
Case Study #1

CANONIZADO, MICHAEL XAVIER, E..

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.

Sample dataset:

ID	Branch_Location	Order_Type	Items_Ordered	Order_Value	Payment_Method	Is_Canceled
01	Albay	Dine-in	3	450	g cash	False
02	Daraga	Take-out	2	280	G-Cash	True
03	Albay	Delivery	5	820	credit card	False
04	SM	Dine-in	1	150	Cash	False
05	Daraga	Delivery	4	600	G cash	False

1) Data Selection

Selected columns:

- ID
- Order_Type
- Items_Ordered
- Order_Value
- Payment_Method
- Customer_Rating
- Is_Canceled

ID	Order_Type	Items_Ordered	Order_Value	Payment_Method	Is_Canceled
01	Dine-in	3	450	g cash	False
02	Take-out	2	280	G-Cash	True
03	Delivery	5	820	credit card	False
04	Dine-in	1	150	Cash	False
05	Delivery	4	600	G cash	False

2) Data Preprocessing

Actions:

- Check for missing values: None
- Make Payment_Method format consistent
- Remove canceled orders

ID	Order_Type	Items_Ordered	Order_Value	Payment_Method	Is_Canceled
01	Dine-in	3	450	g-cash	False
03	Delivery	5	820	credit-card	False
04	Dine-in	1	150	cash	False
05	Delivery	4	600	g-cash	False

3) Data Transformation

Actions:

- Categorize Order_Value: < 300 = LOW, 300 - 600 = MEDIUM, 600+ = HIGH

ID	Order_Type	Items_Ordered	Order_Value	Payment_Method	Is_Canceled	Value_Category
01	Dine-in	3	450	g-cash	False	MEDIUM
03	Delivery	5	820	credit-card	False	HIGH
04	Dine-in	1	150	cash	False	LOW
05	Delivery	4	600	g-cash	False	MEDIUM

4) Data Mining

- Descriptive (Association Rule Discovery)
  - {Value_Category = HIGH or MEDIUM} -> {Payment_Method = g-cash or credit card}
  - People who have a larger bill tend to pay virtually rather than cash
- Descriptive (Summarization)
  - People who order online (Order_Type = Delivery) tend to buy a lot more items and have a Value_Category = HIGH or MEDIUM. Also, if they spend a lot, they are less likely to pay using cash.
- Predictive (Classification)
  - More rows can be added and the model can predict their payment method using the Order_Value column and Order_Type.

5) Pattern Evaluation

- Other datasets can be analyzed to see if the patterns : {Value_Category = HIGH or MEDIUM} -> {Payment_Method = g-cash or credit card} is true.
- Another reliability improvement can be made when we examine other datasets and
- check if the pattern: Order_Type = Delivery -> Value_Category = HIGH or MEDIUM

6) Knowledge Presentation

- Present results in bar charts or pie charts to visually show the patterns
- "People who order online are more likely to buy in bulk"
- "People who buy many items are more likely to pay with virtual money"

7) Knowledge Utilization

- Launch promotions for digital customers so they can buy more items
- Add more payment options for digital customers

8) Feedback

- Monitor the promotions to see if digital buyers are actually availing them
- After a while, rerun the process on new datasets to see if the patterns still exist