REVA UNIVERSITY

Established as per the Section 2(f) of the UGC Act, 1956
Approved by AICTE, COA and BCI, New Delhi

IEEE BANGALORE SECTION
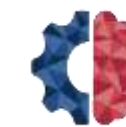
IEEE REVA University Student Branch

Product Affinity Analysis to Increase Sales using Machine Learning

Sharon Joseph(125)
REVA Academy for Corporate Excellence – RACE
REVA University
Bangalore, Karnataka 560064
9945515070
Sharon.ba06@reva.edu.in
Mithun D J, Dr. Rashmi Agarwal

AGENDA

- Introduction
- Literature Review
- Problem Statement
- Proposed Solution
- Objectives
- Methodology
- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment
- Results and Insights
- Conclusion and Future Scope
- References

INTRODUCTION

- Aptean is a product-based company that provides mission-critical, industry-specific software owning an exclusive range of high-end products that serve various domains.

- Like any other company, traditional methods like publishing about the products on the company website, digital marketing, advertisements, and setting up campaigns are followed to sell products.

- However, the organization wanted to explore new ways to increase sales to the existing customer base, thereby increasing the revenue for the organization.

10,000+ Customers

75+ Products

6+ Industry Segments

LITERATURE REVIEW

| Paper Title	Author	Year	Source	Summary	Research Gap
A Recommender System for the Upselling of Telecommunications Products	Navin Dookeram, Zahira Hosein and Patrick Hosein	2022	International Conference on Advanced Communications Technology	This paper focuses on a binary classification framework for predicting the successful upsell of products and services, using data from a telecommunications provider	Data specific to telecommunications industry and only binary classification was considered for analysis
A SVM Ensemble Learning Method Using Tensor Data: An Application to Cross Selling Recommendation	Zhen-Yu Chen, Zhi-Ping Fan and Minghe Sun	2015	International Conference on Service Systems and Service Management	This paper uses a SVM ensemble learning method to propose for classification using tensor data.	Tensor data is used to build the model. SVM is the only methodology focussed on.
Affinity Analysis and Association Rule Mining using Apriori Algorithm in Market Basket Analysis	R. Karthiyayini and Dr. R. Balasubramanian	2016	International Journal of Advanced Research in Computer Science and Software Engineering	The purpose of this analysis is to use Apriori Algorithm in Market Basket Analysis to generate a set of rules that relate two or more products together where lift is above 1.	Apriori algorithm is learnt using the ARules package in R.
Association Rule – Extracting Knowledge Using Market Basket Analysis	Raorane A.A. Kulkarni R.V. and Jitkar B.D.	2012	Research Journal of Recent Sciences	The objective of this paper is to analyze the data exploiting the consumer behavior and employing association rules using Market Basket Analysis to prove its worth over the conventional methodologies.	This study focuses on product placements in supermarkets using Market Basket Analysis.
Market Basket Analysis & Recommendation System Using Association Rules	Shruthi Gurudath	2020	Research gate publication	The goal of this project is to use anonymized data from customers' transactional orders to focus on descriptive analysis of customer purchase patterns, items purchased together, and units purchased frequently from the store to facilitate reordering and maintaining adequate product stock.	Only Market basket analysis is used, and author feels that project can be improved by implementing new and advanced mining algorithms for better performance and fast results.
Market Basket Analysis based on Apriori and CART	Liyuan Wang, Jianqin Sun	2019	International Conference on Education Technology, Management and Humanities Science	The paper which uses Apriori algorithm to find out the data of shopping basket from the massive data of consumers reveals the relationship between the purchased goods, and subsequently applies the association rules and CART decision tree algorithm to reveal the characteristics of the customer group and the target customers classification	Mainly focussed on Market basket analysis using Apriori and CART algorithm to find out the association rules.
Using Data Mining to Accelerate Cross-Selling	Hewen Tang, Zengfang Yang, Pingzhen Zhang and Honglin Yan	2008	International Seminar on Business and Information Management	In this paper, we mainly propose a method of data mining in Excel with an add-in of XLMiner to accelerate cross-selling.	Traditional tool MS excel with XLMiner add-in has been used in this study.

PROBLEM STATEMENT

- Aptean has launched a new growth horizon called, "**Operation 10^X** " and the main goal of this initiative is to drive 10% organic revenue growth each year and become a $ 1 billion revenue company by the end of the next 5 years.

- **Goal**: To Increase sales through product cross-selling and up-selling and identify the right customers to whom these products can be sold.

PROPOSED SOLUTION
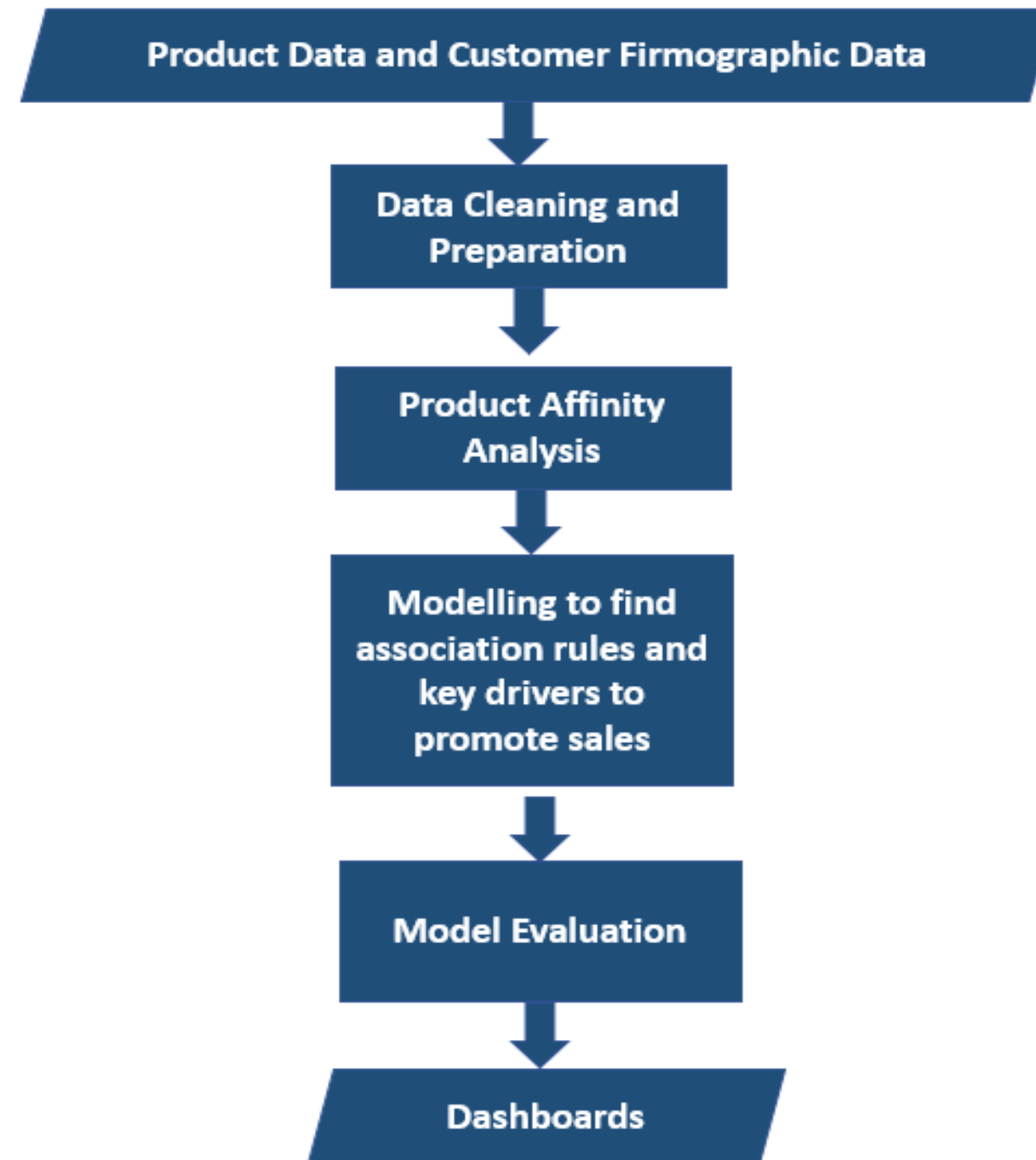
OBJECTIVES

- Three major objectives are,

To understand the purchasing pattern of products from the product sales transaction data.

To study and profile customers based on their purchase behavior.

To recommend and suggest products to customers, thereby increasing sales opportunities.

METHODOLOGY
CRISP-DM Framework

Business Understanding
Goal: To Increase Sales and Revenue

Data Understanding
Product Transaction and Customer Firmographic data collected from Salesforce

Data Preparation
Data Masking, Truth Table, Variable analysis

Modeling
MBA-Apriori, TURF, Decision Tree-CART, KNN, Deep Learning- Multilayer Perceptron

Evaluation
Model evaluation to get optimal Sales

Deployment
Present the outcome to Management

BUSINESS UNDERSTANDING

DATA UNDERSTANDING
DATA COLLECTION

- Initial Data Collected from Organization's internal Salesforce system where Sales transactions and Customer firmographics data are stored in real-time.

- Both the raw datasets have been merged to form one combined dataset for the further scope of analysis. This data is captured and stored in excel format.

DATA UNDERSTANDING
DATA DICTIONARY

Features	Description
Account Name	Name of the Customer in masked format
Order Type	The type of order/transaction
Stage	The Sales Stage
Created Date	Date when the order was created
Close Date	Date when the order was closed
Age	Time taken to close the order (in days)
Year 1 ARR	Financial Data which is masked
Total Software Booking Amount PE	
Services Total	
License Total	
ACV Sub Term Amount	
Product Name	Name of the Product
Region	Product selling region
Product Type	Category of the product
Customer Region	Product buying region
Ownership	Type of ownership of business
Industry	The type of Industry buying the product
Total Revenue	Customers Total annual revenue in masked format

DATA PREPARATION

- Data Selection

- Data Integration

- Data Cleaning

- Data Formatting

- Feature Engineering

DESCRIPTIVE ANALYTICS

- Pie chart A shows 52% of customers are from the Americas region.
- Pie chart B shows 94% of customers are private owners of the business.
- Pie chart C depicts that 52% of customers have purchased products from the SCM and Others categories.
- Pie chart D shows that only 75% of the opportunities or orders were successfully closed.

DESCRIPTIVE ANALYTICS

No of Purchases by Products

A

No of Products by Product Category

B

No of Products by Region

C
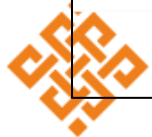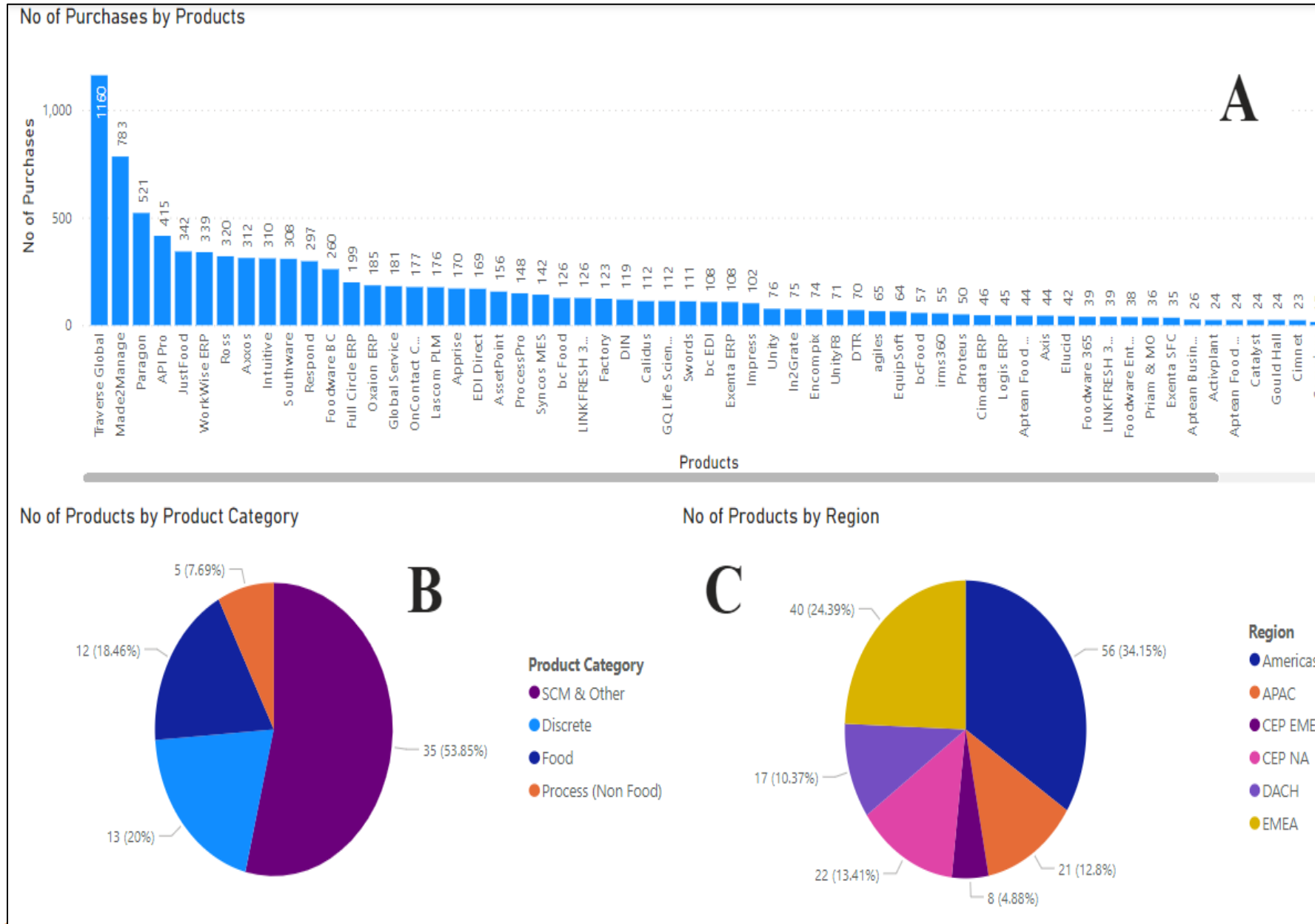
- Bar chart A shows the top-selling products in descending order.

- Pie chart B shows that 35 products fall under the product category SCM and others.

- Pie chart C shows the number of products by region and there are 56 products sold in the Americas region, thereby making the Americas the major selling site across the globe.

MODELING

Overview of the dataflow into the Machine Learning Model
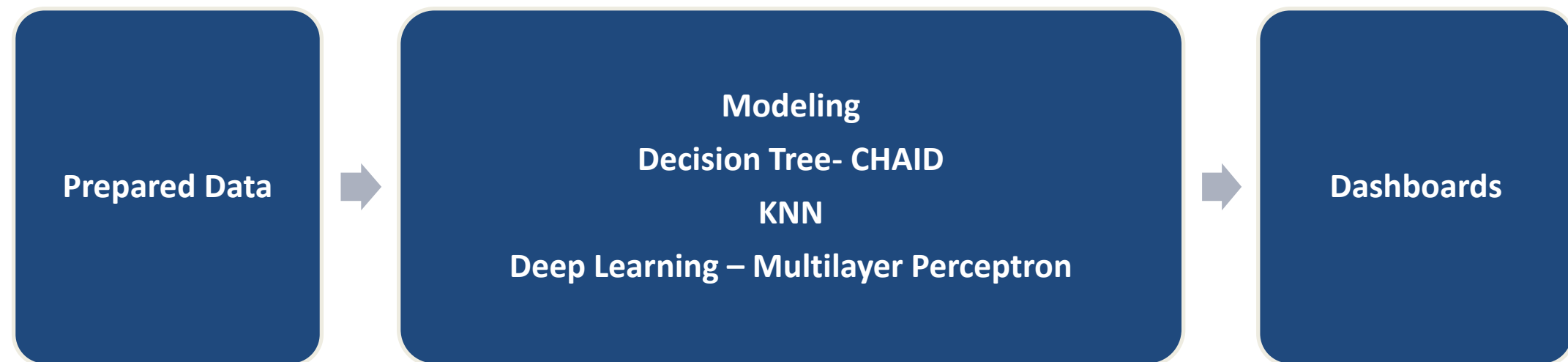
TURF ANALYSIS

		"Best TURF Results"			
	Statistics				
Features	Size of group	"Reach"	"% of Cases"	"Frequency"	"% of Responses"
ADDED: TraverseGlobal	1	1160	13.2	1160	16.1
ADDED: Made2Manage KEPT: TraverseGlobal	2	1941	22.1	1943	27.0
ADDED: Paragon KEPT: Made2Manage, TraverseGlobal	3	2462	28.0	2464	34.3
ADDED: APIPro KEPT: Made2Manage, Paragon, TraverseGlobal	4	2877	32.7	2879	40.0
ADDED: JustFood KEPT: APIPro, Made2Manage, Paragon, TraverseGlobal	5	3216	36.5	3221	44.8
ADDED: WorkWiseERP KEPT: APIPro, JustFood, Made2Manage, Paragon, TraverseGlobal	6	3549	40.3	3560	49.5

MARKET BASKET ANALYSIS
Apriori algorithm

Consequent	Antecedent	Instances	Support %	Confidence %	Rule Support %	Lift	Deployability
EDI Direct	Full Circle ERP	81	13.61	98.77	13.45	7.17	0.17
Foodware BC	Foodware 365	35	5.88	94.29	5.55	8.01	0.34
Oxaion ERP	Syncos MES	28	4.71	92.86	4.37	17.82	0.34
OnContact CRM	WorkWise ERP	81	13.61	92.59	12.61	5.35	1.01
bc Food	bc EDI	45	7.56	62.22	4.71	6.38	2.86
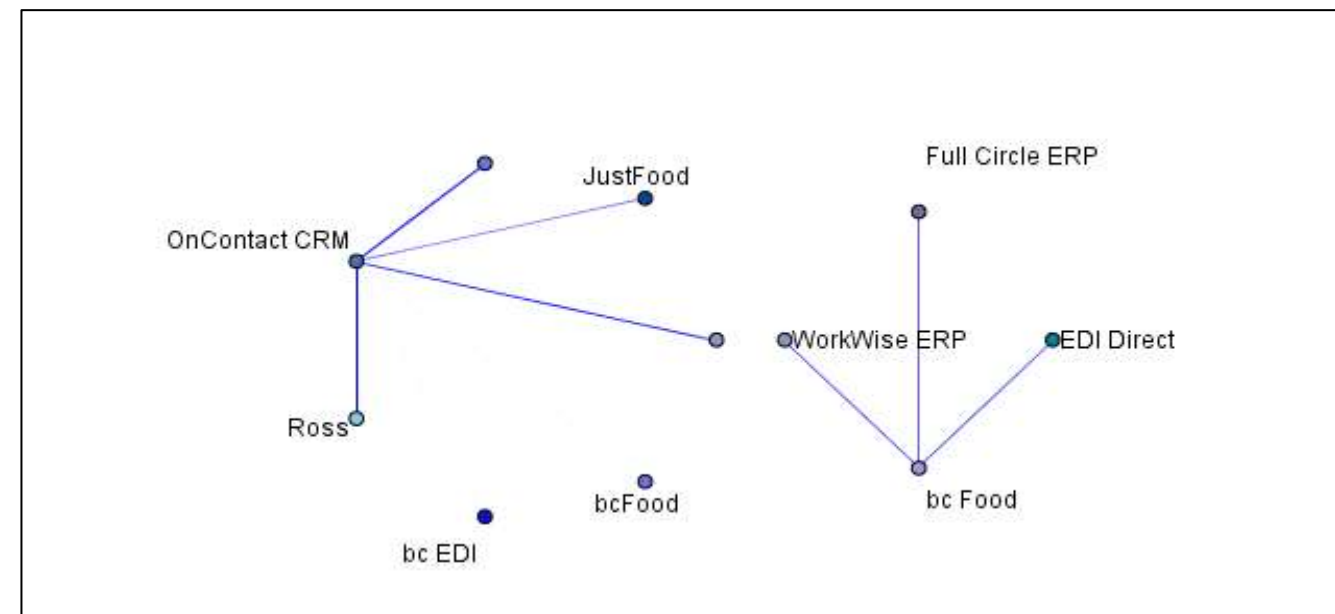
DECISION TREE
CHAID Method

DV-EDI+FullCircle

Node 0

Category	%	n
■ 0.000	60.000	120
■ 1.000	40.000	80
Total	100.000	200

Product Region
Adj. P-value=0.000, Chi-square=62.745, df=1

APAC; CEP EMEA; CEP NA; DACH; EMEA

Node 1

Category	%	n
■ 0.000	100.000	64
■ 1.000	0.000	0
Total	32.000	64

Americas

Node 2

Category	%	n
■ 0.000	41.176	56
■ 1.000	58.824	80
Total	68.000	136

Product Type
Adj. P-value=0.000, Chi-square=37.494, df=1

Discrete; Food; Process (Non Food)

Node 3

Category	%	n
■ 0.000	100.000	22
■ 1.000	0.000	0
Total	11.000	22

SCM & Other

Node 4

Category	%	n
■ 0.000	29.825	34
■ 1.000	70.175	80
Total	57.000	114

Ownership
Adj. P-value=0.013, Chi-square=9.754, df=1

Nonprofit; Public

Node 5

Category	%	n
■ 0.000	100.000	4
■ 1.000	0.000	0
Total	2.000	4

Partnership; Private

Node 6

Category	%	n
■ 0.000	27.273	30
■ 1.000	72.727	80
Total	55.000	110

Predictor Importance
Target: DV-EDI+FullCircle

Product Region

Product Type

Ownership

0.0   0.2   0.4   0.6   0.8   1.0

Least Important   Most Important

KNN

Predictor Space

Built Model: 3 selected predictors, K = 3

Select points to use as focal records

This chart is a lower-dimensional projection of the predictor space, which contains a total of 5 predictors.

DEEP LEARNING
Multilayer Perceptron
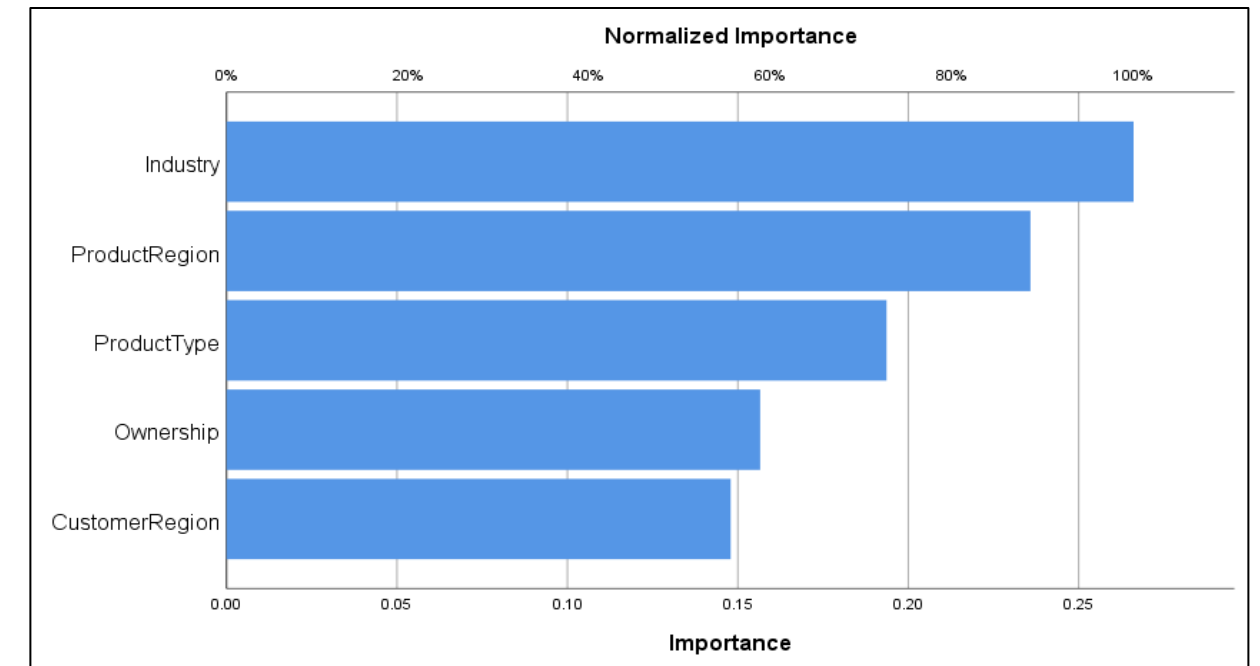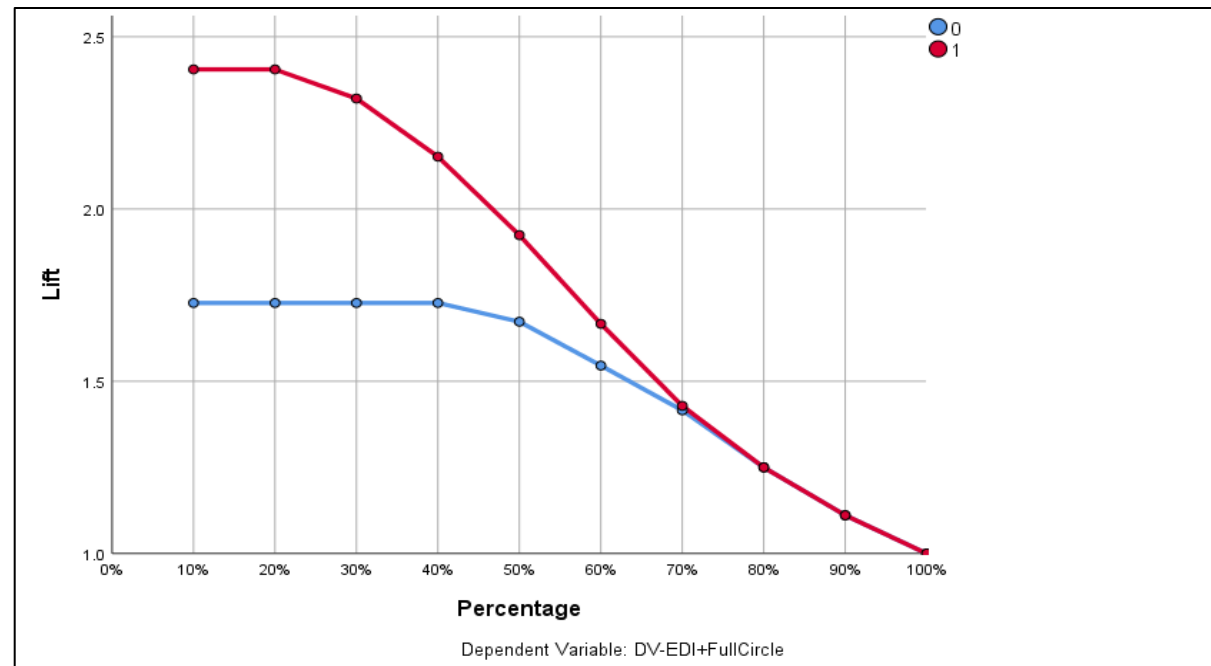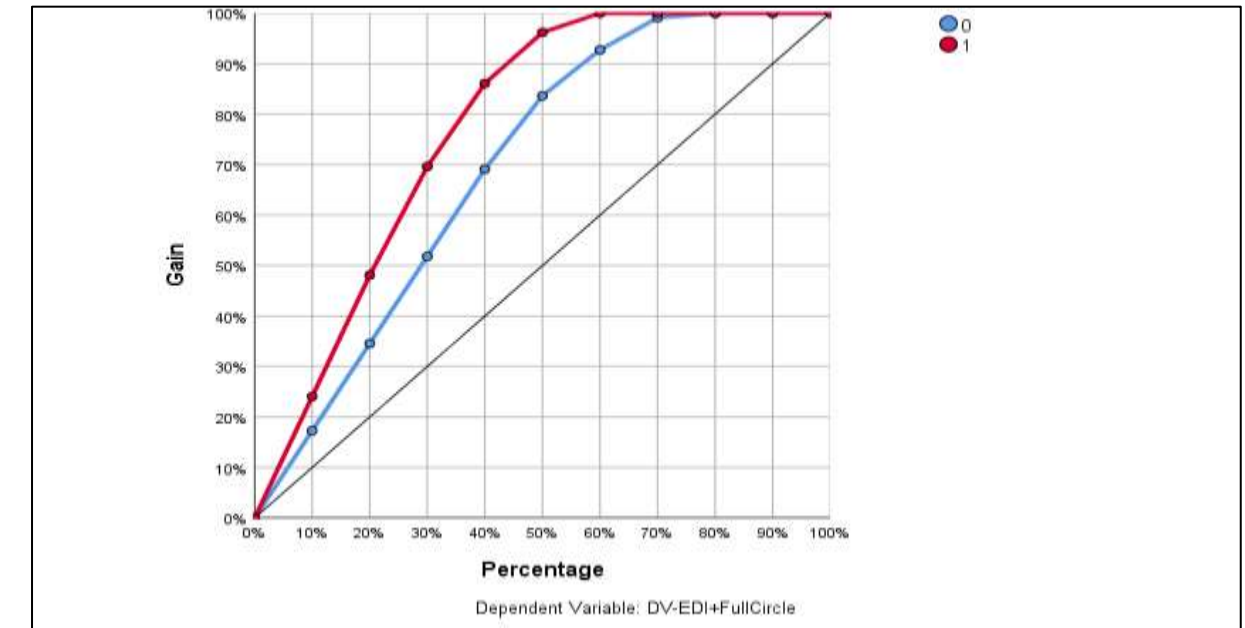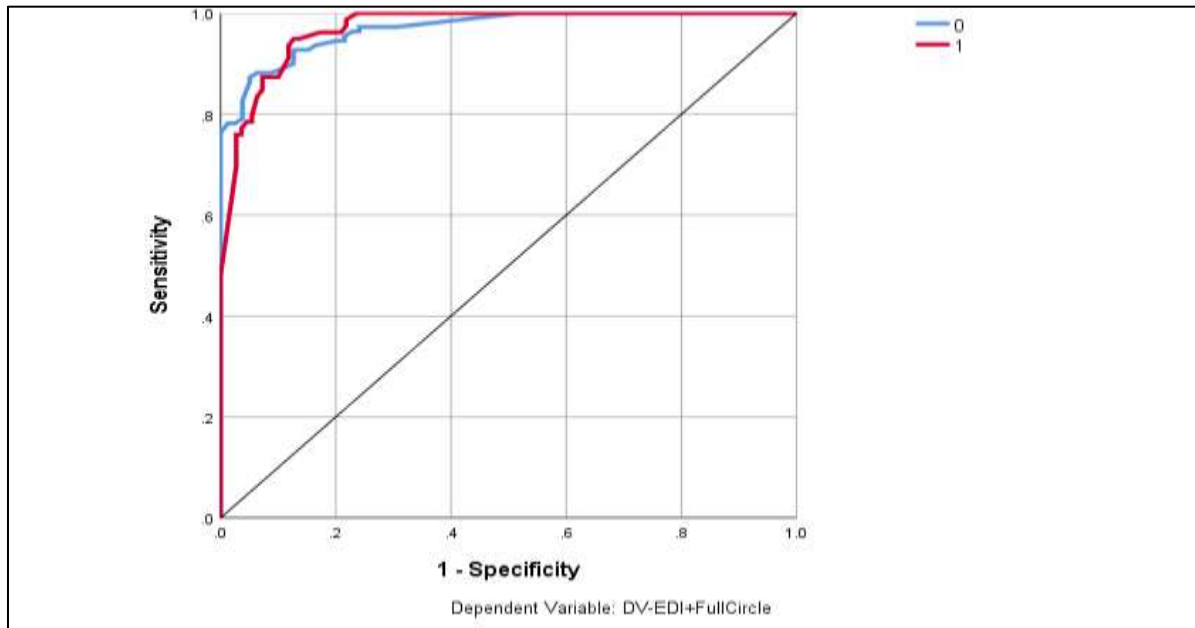
EVALUATION

- The models are evaluated using the confusion matrix and the model performance is calculated using the True Positive (TP), False Positive (FP), True Negative (TN) and False Negative (FN) values.

- The Decision tree using CHAID has a higher accuracy score of 92% and a Precision value of 94%, proving better performance.

Model Performance metrics				
Models	Precision	Recall	F1-score	Accuracy
Decision Tree	94%	84%	89%	92%
KNN	89%	71%	79%	85%
MLP- Test	77%	100%	87%	89%
MLP-Train	87%	94%	90%	91%

DEPLOYMENT

Pick the best model to predict the sales opportunities

Create a Dashboard in Power BI with the findings of the model

Present the Dashboard to management and relevant stakeholders

Get feedback and improvise the model

Share the results with the Sales team and relevant stakeholders to implement the model.
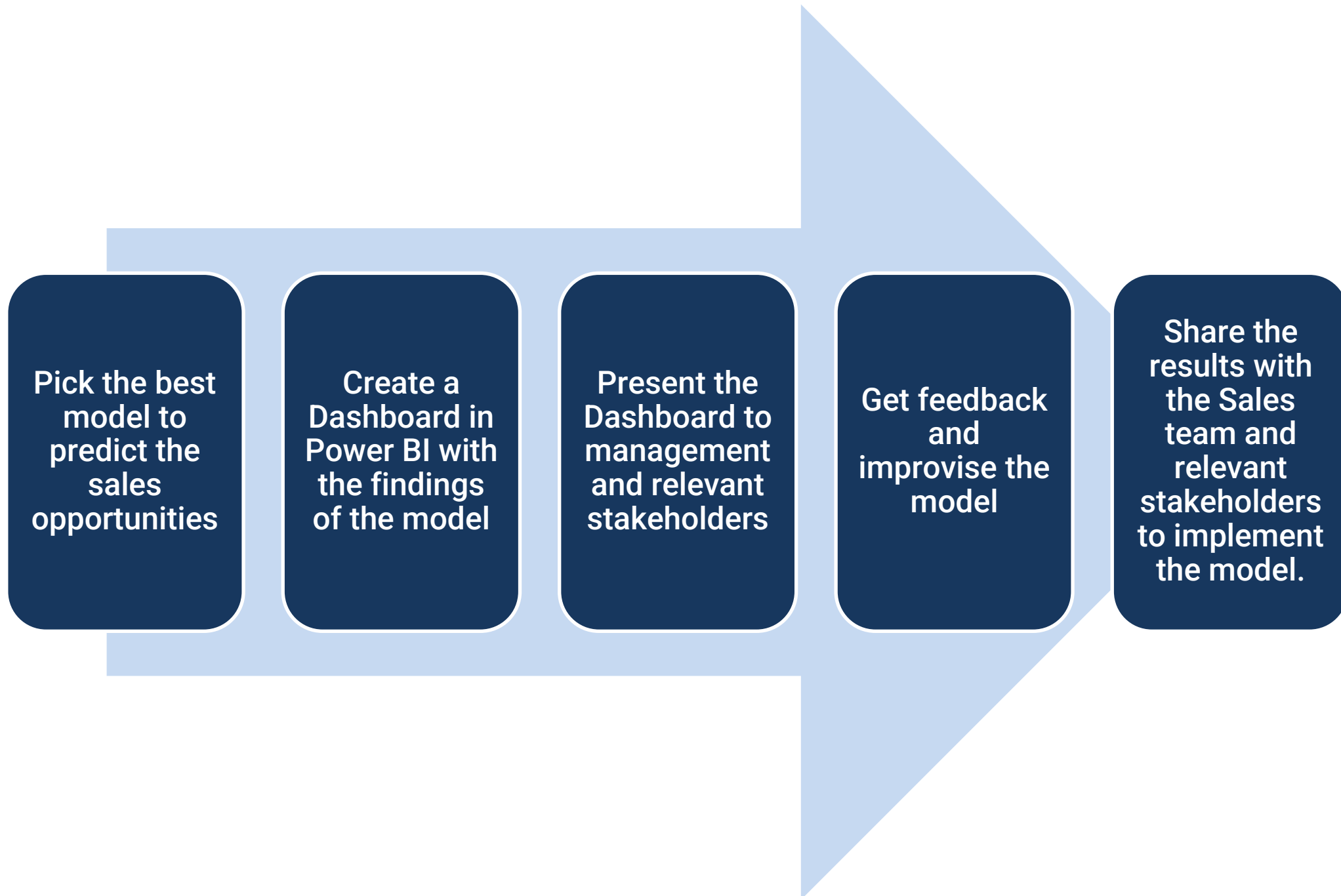
RESULTS AND INSIGHTS

- The most efficient model used in this study is the Decision tree using the CHAID method, because of its high accuracy and precision value.

Product A	Product B	Instances	Product Category
Full Circle ERP	EDI Direct	80	SCM and Other
Foodware 365	Foodware BC	33	Food
Syncos MES	Oxaion ERP	26	SCM and Other
WorkWise ERP	OnContact CRM	75	SCM and Other
bc EDI	bc Food	35	Food
Ross	Factory	14	Process

Key drivers influencing product sales as shown in the decision tree are –

- **Product Region** – Selling region of the product
- **Product Type** – To which category of product it belongs
- **Ownership** – The type of ownership of the customer company

CONCLUSION AND FUTURE SCOPE

Conclusion:

- Different modelling techniques have been evaluated to find out the key drivers responsible for cross-selling products.

- We could profile customers that belong to different categories based on these key drivers and propose that for new customers who belong to any of these categories, such products could be sold, thereby increasing sales opportunities.

Future Scope:

- This project does not cover the cost and financial analysis, if the financial data could be used for analysis, we could probably recommend the best possible products for upselling or cross-selling thereby increasing sales.

- A similar analysis can be used to model other combinations of data in which more than 2 products are sold together.

REFERENCES

[1] Aman Kharwal. (2021). Classification Report. https://thecleverprogrammer.com/2021/07/07/classification-report-in-machine-learning/#:~:text=A%20classification%20report%20is%20a,this%20article%20is%20for%20you.

[2] Chen, Z. Y., Fan, Z. P., & Sun, M. (2015, July 28). An SVM ensemble learning method using tensor data: An application to cross-selling recommendation. 2015 12th International Conference on Service Systems and Service Management, ICSSSM 2015. https://doi.org/10.1109/ICSSSM.2015.7170282

[3] David Gargaro. (2022). 12-ways-to-increase-sales. https://www.business.com/articles/12-ways-to-increase-sales/

[4] Dookeram, N., Hosein, Z., & Hosein, P. (2022). A Recommender System for the Upselling of Telecommunications Products. International Conference on Advanced Communication Technology, ICACT, 2022-February, 66–72. https://doi.org/10.23919/ICACT53585.2022.9728818

[5] Gupta, T., Karthiyayini, R., Balasubramanian Professor, R., & And, P. (2016a). Affinity and Association Affinity Analysis and Association Rule Mining using Apriori Algorithm in Market Basket Analysis. In International Journal of Advanced Research in Computer Science and Software Engineering (Vol. 6, Issue 10). www.ijarcsse.com

[6] Gupta, T., Karthiyayini, R., Balasubramanian Professor, R., & And, P. (2016b). Affinity and Association Affinity Analysis and Association Rule Mining using Apriori Algorithm in Market Basket Analysis. In International Journal of Advanced Research in Computer Science and Software Engineering (Vol. 6, Issue 10). www.ijarcsse.com

[7] Hewen, T., Zengfang, Y., Pingzhen, Z., & Honglin, Y. (2008). Using data mining to accelerate cross-selling. 2008 International Seminar on Business and Information Management, ISBIM 2008, 1, 283–286. https://doi.org/10.1109/ISBIM.2008.186

[8] Raorane AA, Kulkarni RV, & Jitkar BD. (2012). Association Rule-Extracting Knowledge Using Market Basket Analysis. www.isca.in

[9] Shruthi Gurudath. (2020). Market Basket Analysis & Recommendation System Using Association Rules. https://doi.org/10.13140/RG.2.2.16572.05767

[10] Think Insights. (2022). CRISP-DM – A framework for Data Mining & Analysis. https://thinkinsights.net/data-literacy/crisp-dm/ Page 43 of 45

[11] Turf Analysis. (2019). https://conjointly.com/blog/turf-analysis/

[12] Wikipedia. (2022). Upselling. https://en.wikipedia.org/wiki/Upselling

Thank you

ICSTCEE 2022, REVA University, Bangalore