REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

REVA Academy for Corporate Excellence (RACE)

Product Affinity Analysis using Machine Learning

SHARON JOSEPH

SRN: R19DM053

Date: 27-08-2022

MBA in Business Analytics

Capstone Project Presentation
Year: II

race.reva.edu.in

REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

Agenda

01 Introduction
Background | Current status | Why this study

02 Literature Review
Seminal works | Summary | Research Gap

03 Problem Statement
Business Problem | Analytics Solution

04 Project Objectives
Primary & Secondary Objectives | Expected Outcome

05 Project Methodology
Conceptual Framework | Research Design

06 Business Understanding
Business Context | Monetary Impact

07 Data Understanding
Data Collection | Variables

08 Data Preparation
Pre-processing | Process | Techniques

09 Descriptive Analytics
Univariate | Bivariate | Hypothesis

10 Modeling
Machine Learning | Model Evaluation | Insights

11 Model Deployment
Applications | Demo

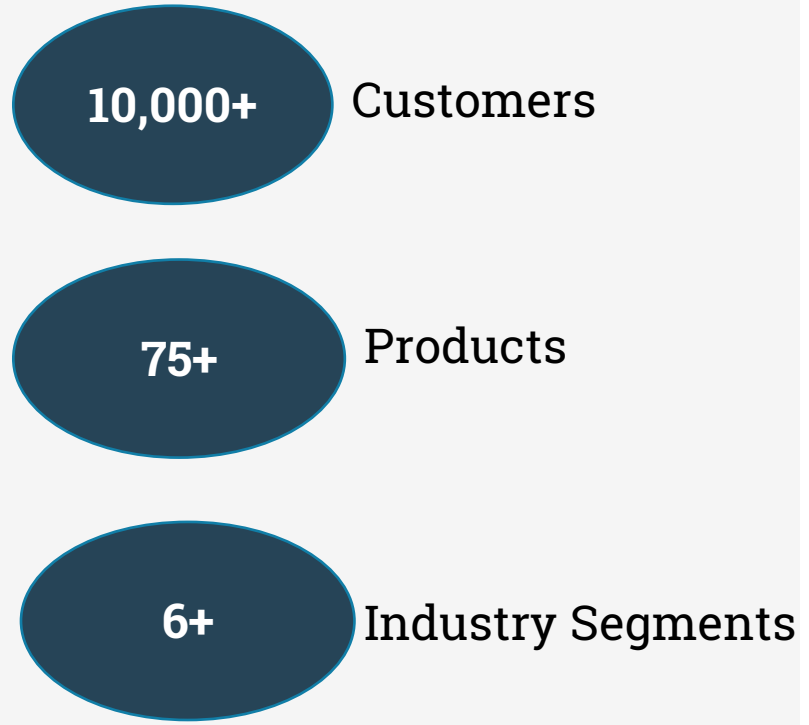
12 Suggestions and Conclusions
Insights | Next Step | Future Scope

13 Annexure
References | Publications | Plagiarism Score

Introduction

Background:

Aptean is a product-based company that provides mission-critical, industry-specific software owning an exclusive range of high-end products that serve in 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 in which sales could be increased to the existing customer base, thereby increasing the revenue for the organization.



Why this study?

Working as a BI analyst, in the Strategy department of my organization, we get requests from several stakeholders to build reports and dashboards so that the executive leadership team can have a birds-eye view on the company statistics. Therefore, rather than the regular dashboarding work, I wanted to include machine learning models to improvise on any existing methodology, that the organization wanted to change. One such area was to improve sales strategies and hence this study.

Literature Review

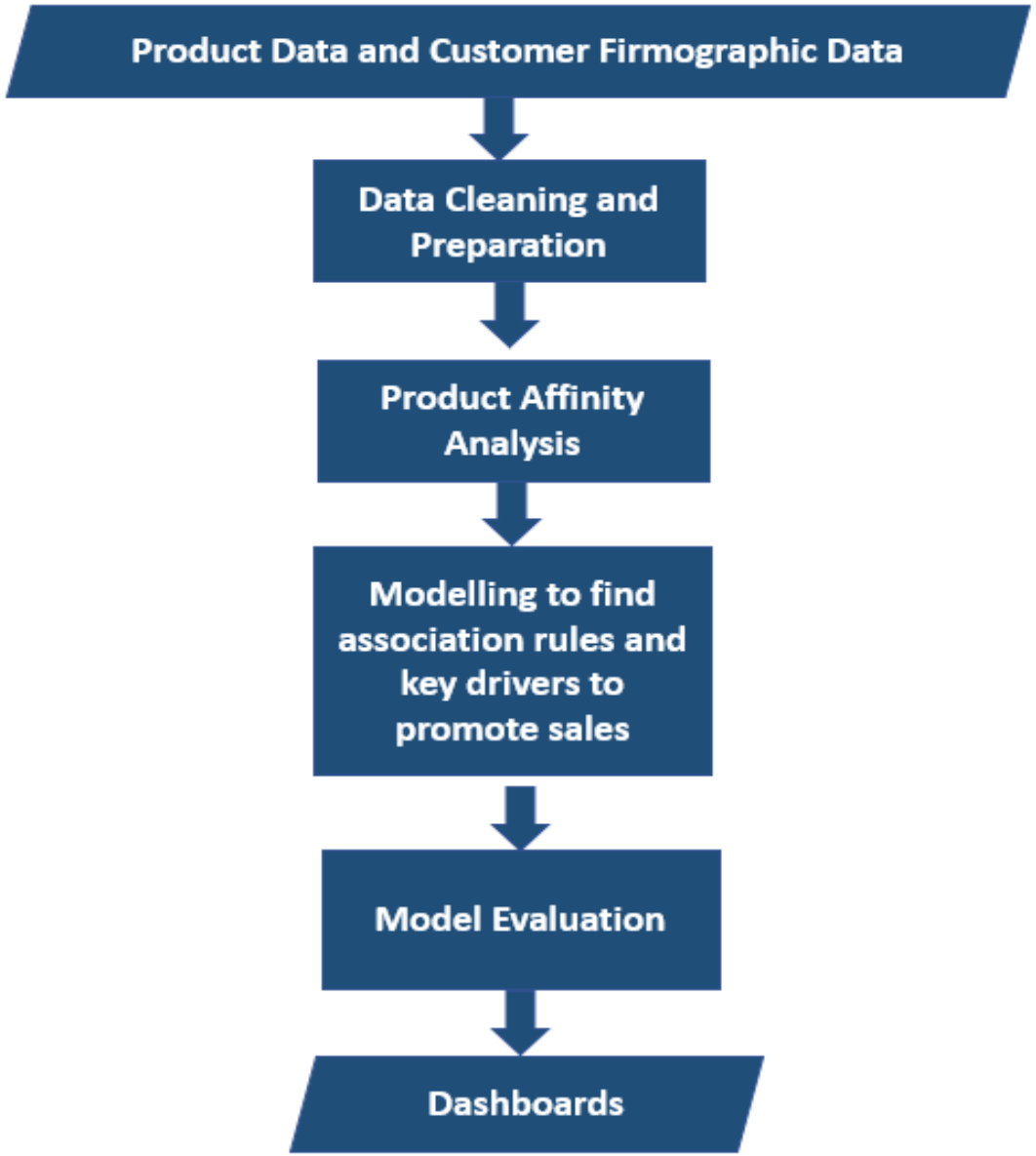
Paper Title	Author	Year	Source	Summary	Research Gap
A Recommender System for the Upselling of Telecommunications Products	Navin Dookeram, Zahira Hosein and Patrick Hosein	2022	International Conference on Advanced Communications Technology	This paper focuses on a binary classification framework for predicting 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 addin has been used in this study.

Problem Statement

Aptean has launched the next 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.

Leverage customer firmographic data and product sales transaction data to build a model to project the likelihood to purchase from our existing customer base and solve the business challenge of achieving and meeting the sales targets of the organization. Increasing sales through product cross-selling and up-selling and identifying the customers to whom products can be sold

Proposed Solution



Project Objectives

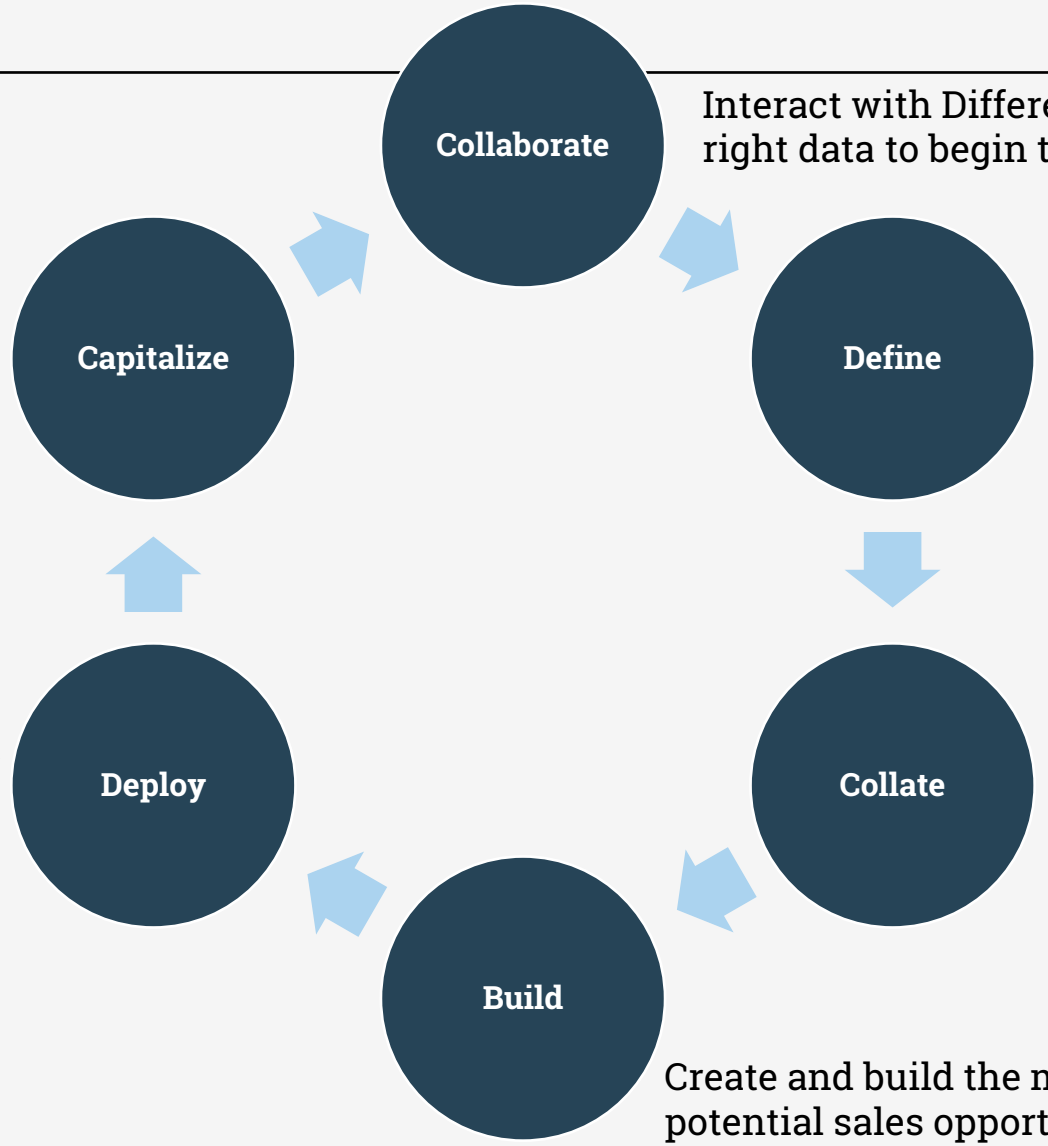
Three major objectives of this study are,

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.

Project Methodology- FLUID



Interact with Different teams to get the right data to begin the analysis

Align to meet the business objectives and goal of increasing product sales

Perform EDA and clean the data to suit the requirements

Create and build the model in order to identify potential sales opportunities

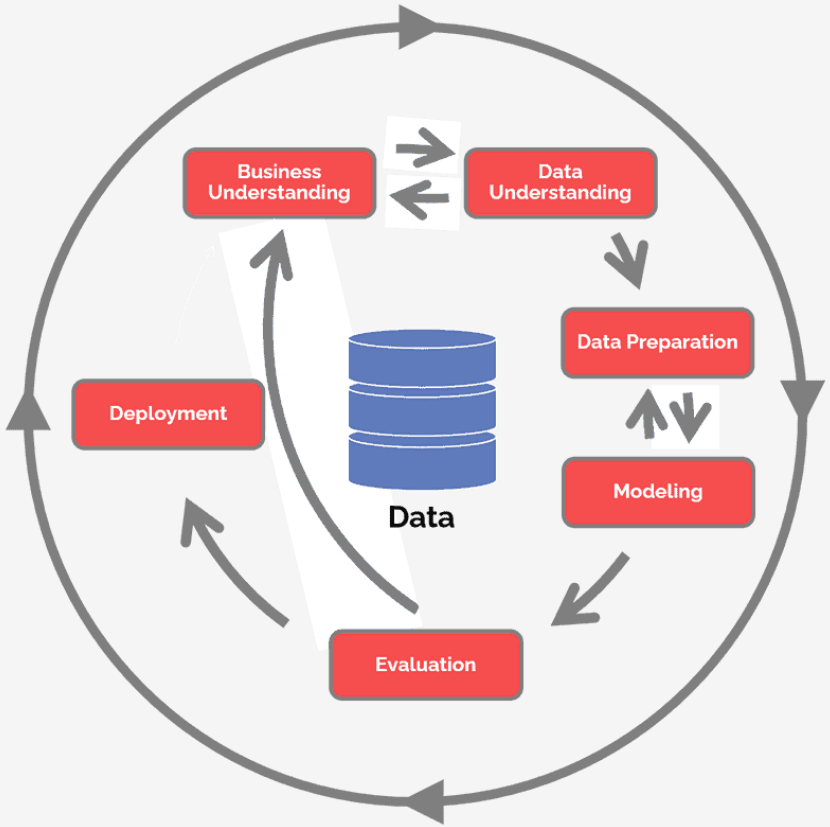
Deploy the model and improvise it for better sales predictions by creating dashboards

Increase Sales and predict potential customers

REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956, Approved by AICTE, New Delhi

Project Methodology- CRISP-DM

CRISP-DM Framework: The Cross-Industry Standard Process for Data Mining methodology is used in the implementation of this project.



Source: CRISP-DM from
https://www.datascience-pm.com/crisp-dm-2/

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

Collaborate to have Business Understanding

7 Stages	Sales Team Activities
Buyer Awareness	The opportunity is reviewed, and the sales team will determine if they accept the meeting.
Connect	The sales team confirms the solution fit and plans the priority, consequences, budget range and timeline and the next meeting is scheduled within 2 weeks.
Educate	Conducts detailed qualifications. Identifies buyer roles, potential catalysts, competitors, and purchase process. Conducts Overview Demo.
Build	Business and fitness assessments. Validates buyer insights with stakeholders. Create an action plan. Prepare buyer scorecard/ Decision criteria. Create a preliminary proposal.
Validate	Proactive project implementation plan discussions. Address customer security requirements. Create a formal proposal/quote.
Confirm	Facilitate reference introductions. Review contract and confirm purchasing process. Negotiate/support cost justifications.
Close	Check with finance teams and prepare the closure plan.

Define

In the define phase , after the collaboration with management and relevant teams, the project goals were identified.

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

Data Understanding

DATA COLLECTION:

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

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

DATA DICTIONARY

These are the features considered in the dataset.

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

REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956, Approved by AICTE, New Delhi

Collate + Data Preparation

1.Data Selection

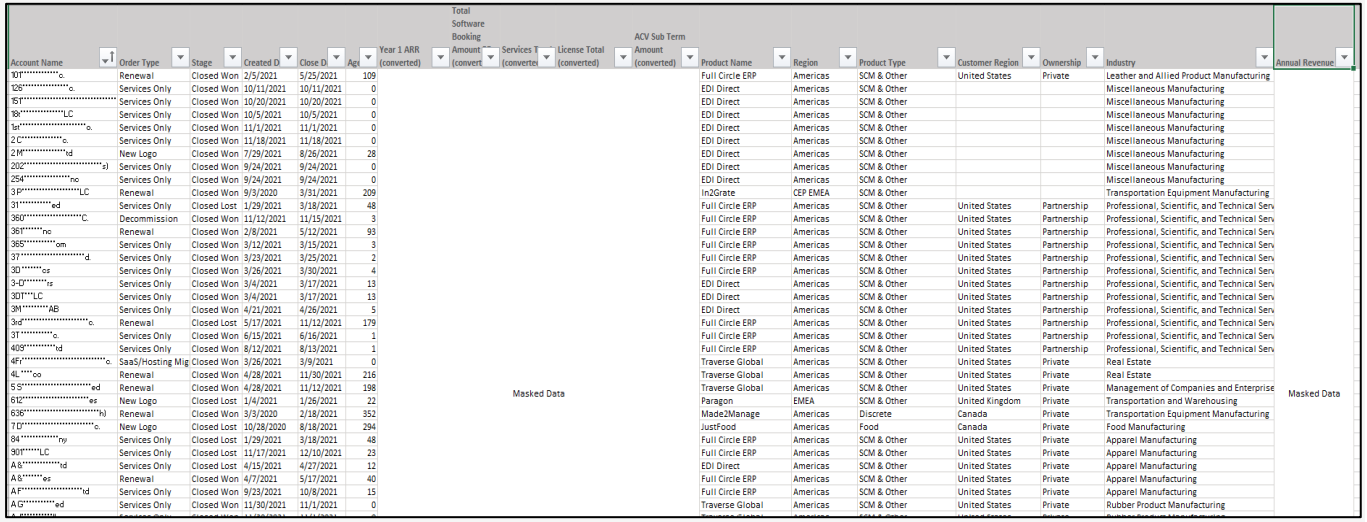
2 Data Integration

3 Data Cleaning

- 3.1 Missing value handling
- 3.2 Handling erroneous/misspelled data
- 3.3 Handling duplicate data

4 Data Formatting

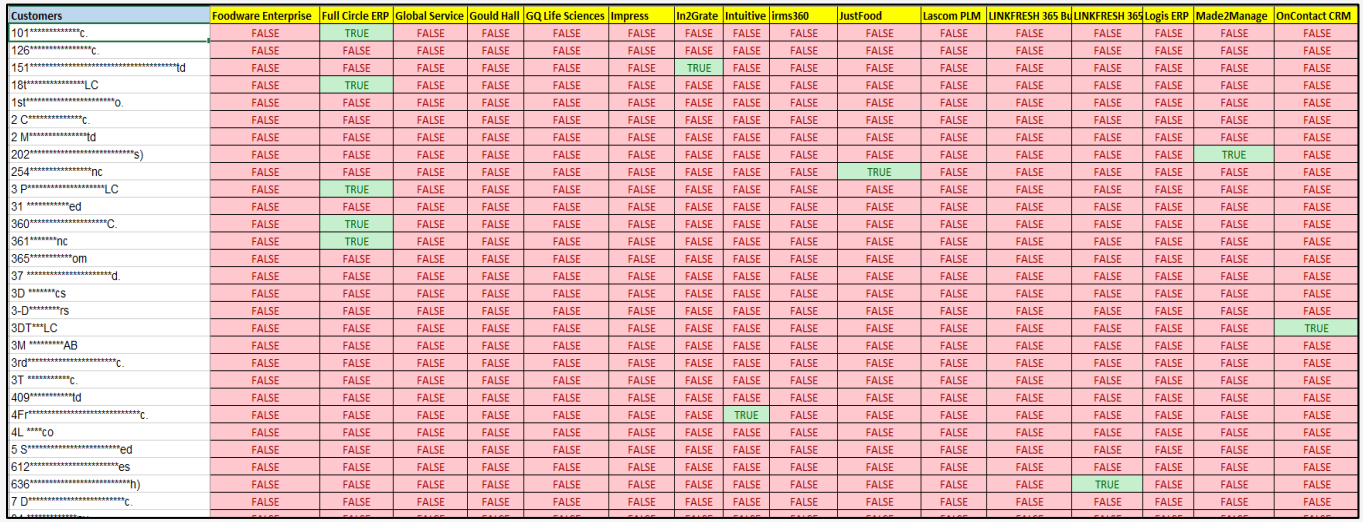
5 Feature Engineering



Account Name	Order Type	Stage	Created D	Close D	Age	Year 1 ARR (converted)	Total Software Booking Amount (convert	Services T (converte	License Total (converted)	ACV Sub Term Amount (converted)	Product Name	Region	Product Type	Customer Region	Ownership	Industry	Annual Revenue
10T**********o.	Renewal	Closed Won	2/5/2021	5/25/2021	109						Full Circle ERP	Americas	SCM & Other	United States	Private	Leather and Allied Product Manufacturing	
126**********c.	Services Only	Closed Won	10/11/2021	10/11/2021	0						EDI Direct	Americas	SCM & Other			Miscellaneous Manufacturing	
151**********	Services Only	Closed Won	10/20/2021	10/20/2021	0						EDI Direct	Americas	SCM & Other			Miscellaneous Manufacturing	
18r**********LC	Services Only	Closed Won	10/5/2021	10/5/2021	0						EDI Direct	Americas	SCM & Other			Miscellaneous Manufacturing	
1st**********o.	Services Only	Closed Won	11/1/2021	11/1/2021	0						EDI Direct	Americas	SCM & Other			Miscellaneous Manufacturing	
2 C**********c.	Services Only	Closed Won	11/18/2021	11/18/2021	0						EDI Direct	Americas	SCM & Other			Miscellaneous Manufacturing	
2 M**********td	New Logo	Closed Won	7/29/2021	8/26/2021	28						EDI Direct	Americas	SCM & Other			Miscellaneous Manufacturing	
202**********s)	Services Only	Closed Won	9/24/2021	9/24/2021	0						EDI Direct	Americas	SCM & Other			Miscellaneous Manufacturing	
254**********nc	Services Only	Closed Won	9/24/2021	9/24/2021	0						EDI Direct	Americas	SCM & Other			Miscellaneous Manufacturing	
3P**********LC	Renewal	Closed Won	9/3/2020	3/31/2021	209						In2Grate	CEP EMEA	SCM & Other			Transportation Equipment Manufacturing	
31**********ed	Services Only	Closed Lost	1/29/2021	3/18/2021	48						Full Circle ERP	Americas	SCM & Other	United States	Partnership	Professional, Scientific, and Technical Serv	
360**********C.	Decommission	Closed Won	11/12/2021	11/15/2021	3						Full Circle ERP	Americas	SCM & Other	United States	Partnership	Professional, Scientific, and Technical Serv	
361*******nc	Renewal	Closed Won	2/8/2021	5/12/2021	93						Full Circle ERP	Americas	SCM & Other	United States	Partnership	Professional, Scientific, and Technical Serv	
365**********om	Services Only	Closed Won	3/12/2021	3/15/2021	3						Full Circle ERP	Americas	SCM & Other	United States	Partnership	Professional, Scientific, and Technical Serv	
37**********d.	Services Only	Closed Won	3/23/2021	3/25/2021	2						Full Circle ERP	Americas	SCM & Other	United States	Partnership	Professional, Scientific, and Technical Serv	
3D*******cs	Services Only	Closed Won	3/26/2021	3/30/2021	4						Full Circle ERP	Americas	SCM & Other	United States	Partnership	Professional, Scientific, and Technical Serv	
3-D********rs	Services Only	Closed Won	3/4/2021	3/17/2021	13						EDI Direct	Americas	SCM & Other	United States	Partnership	Professional, Scientific, and Technical Serv	
3DT***LC	Services Only	Closed Won	3/4/2021	3/17/2021	13						EDI Direct	Americas	SCM & Other	United States	Partnership	Professional, Scientific, and Technical Serv	
3M *********AB	Services Only	Closed Won	4/21/2021	4/26/2021	5						EDI Direct	Americas	SCM & Other	United States	Partnership	Professional, Scientific, and Technical Serv	
3rd**********c.	Renewal	Closed Lost	5/17/2021	11/12/2021	179						Full Circle ERP	Americas	SCM & Other	United States	Partnership	Professional, Scientific, and Technical Serv	
3T**********c.	Services Only	Closed Won	6/15/2021	6/16/2021	1						Full Circle ERP	Americas	SCM & Other	United States	Partnership	Professional, Scientific, and Technical Serv	
409**********td	Services Only	Closed Won	8/12/2021	8/13/2021	1						Full Circle ERP	Americas	SCM & Other	United States	Partnership	Professional, Scientific, and Technical Serv	
4F**********c.	SaaS/Hosting Mig	Closed Won	3/26/2021	3/9/2021	0						Traverse Global	Americas	SCM & Other	United States	Private	Real Estate	
4L ****co	Renewal	Closed Won	4/28/2021	11/30/2021	216						Traverse Global	Americas	SCM & Other	United States	Private	Real Estate	
5 S**********ed	Renewal	Closed Won	4/28/2021	11/12/2021	198						Traverse Global	Americas	SCM & Other	United States	Private	Management of Companies and Enterprise	
612**********es	New Logo	Closed Lost	1/4/2021	1/26/2021	22						Paragon	EMEA	SCM & Other	United Kingdom	Private	Transportation and Warehousing	
636**********h)	Renewal	Closed Won	3/3/2020	2/18/2021	352						Made2Manage	Americas	Discrete	Canada	Private	Transportation Equipment Manufacturing	
7 D**********c.	New Logo	Closed Lost	10/28/2020	8/18/2021	294						JustFood	Americas	Food	Canada	Private	Food Manufacturing	
84**********ny	Services Only	Closed Lost	1/29/2021	3/18/2021	48						Full Circle ERP	Americas	SCM & Other	United States	Private	Apparel Manufacturing	
90T*****LC	Services Only	Closed Lost	11/17/2021	12/10/2021	23						Full Circle ERP	Americas	SCM & Other	United States	Private	Apparel Manufacturing	
A 8**********td	Services Only	Closed Lost	4/15/2021	4/27/2021	12						EDI Direct	Americas	SCM & Other	United States	Private	Apparel Manufacturing	
A 8*******es	Renewal	Closed Won	4/7/2021	5/17/2021	40						Full Circle ERP	Americas	SCM & Other	United States	Private	Apparel Manufacturing	
A F**********td	Services Only	Closed Won	9/23/2021	10/8/2021	15						Full Circle ERP	Americas	SCM & Other	United States	Private	Apparel Manufacturing	
A G**********ed	Services Only	Closed Won	11/30/2021	11/1/2021	0						Traverse Global	Americas	SCM & Other	United States	Private	Rubber Product Manufacturing	

Masked Data

Masked Data



Customers	Foodware Enterprise	Full Circle ERP	Global Service	Gould Hall	GQ Life Sciences	Impress	In2Grate	Intuitive	irms360	JustFood	Lascom PLM	LINKFRESH 365 Bu	LINKFRESH 365	Logis ERP	Made2Manage	OnContact CRM
101**********c.	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
126**********c.	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
151**********td	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
18t**********LC	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
1st**********o.	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
2 C**********c.	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
2 M**********td	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
202**********s)	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE
254**********nc	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
3 P**********LC	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
31 **********ed	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
360**********C.	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
361*******nc	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
365**********om	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
37 **********d.	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
3D *******cs	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
3-D********rs	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
3DT***LC	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE
3M *********AB	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
3rd**********c.	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
3T **********c.	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
409**********td	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
4Fr**********c.	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
4L ****co	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
5 S**********ed	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
612**********es	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE
636**********h)	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE
7 D**********c.	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE

Descriptive Analytics

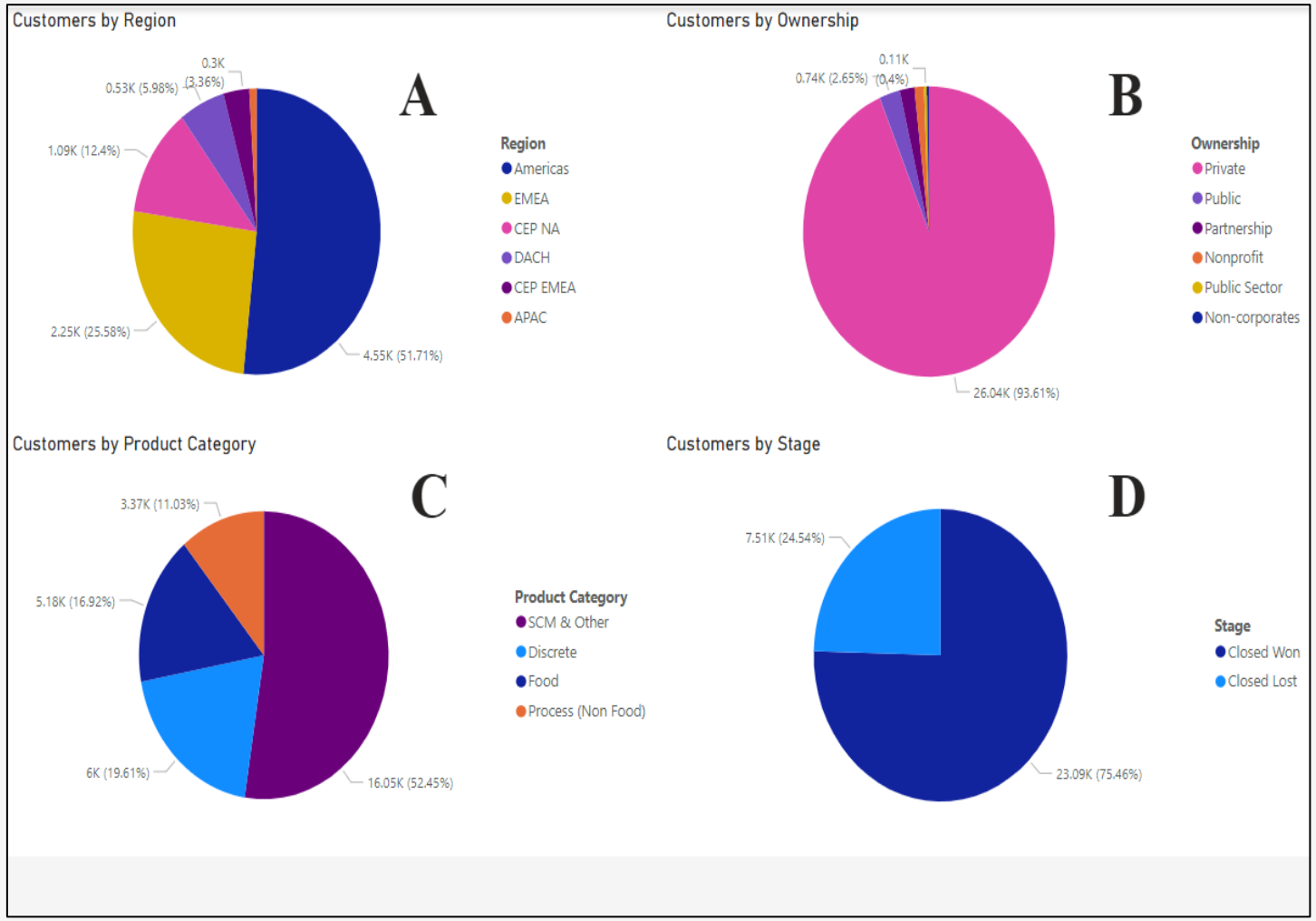


Figure has 4 pie charts, A, B, C, and D. Here, pie chart A shows the customers by region, and it can be concluded that close to 52% of customers are from the Americas region. Pie chart B shows the customers by ownership, where about 94% of customers are private owners of the business. Pie chart C depicts that 52% of customers have purchased products from the SCM and Others category. Finally , pie chart D shows that out of all the product transactions, only 75% of the opportunities or orders were successfully closed.

Descriptive Analytics

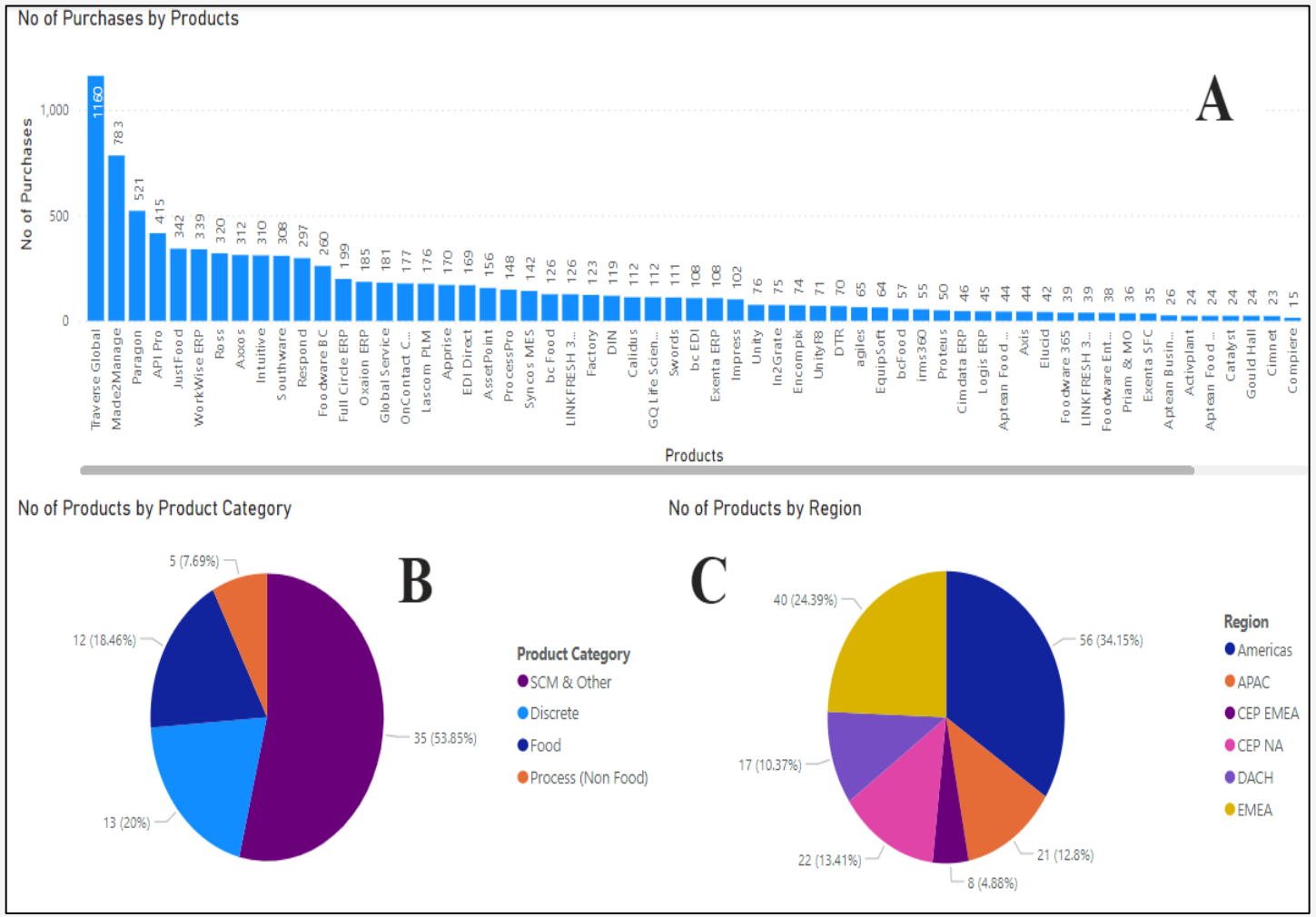
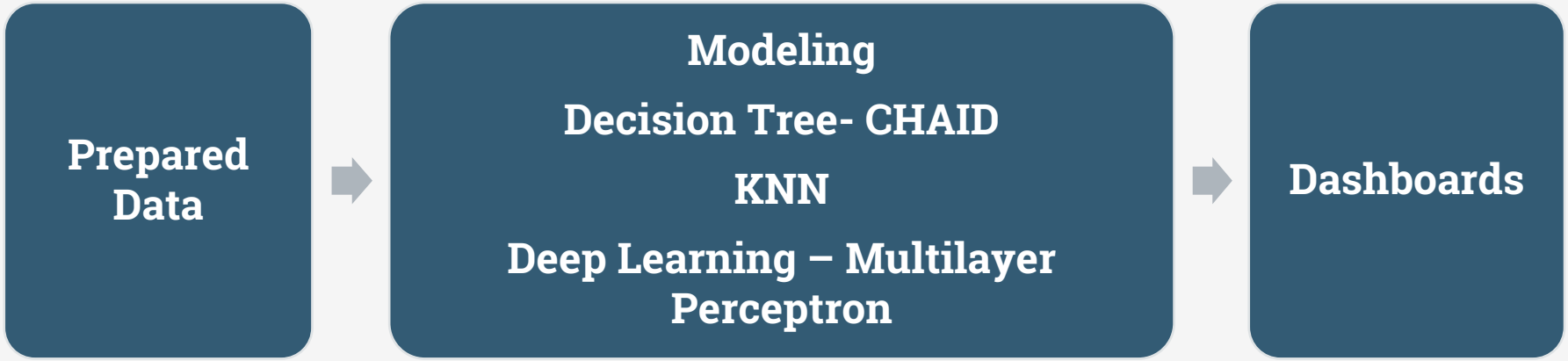


Figure 7.3 has visuals A, B and C. Bar chart A shows the top selling products in the descending order. Traverse Global, Made2Manage and Paragon are the top 3 selling products. Pie chart B shows that 35 products fall under the product category SCM and other. This could be reason why the most transactions have happened for the same category. Pie chart C shows the number of products by region and there are 56 products sold in the Americas region, thereby making Americas the major selling site across the globe

Build + Modeling

Overview of the dataflow into Machine Learning Model



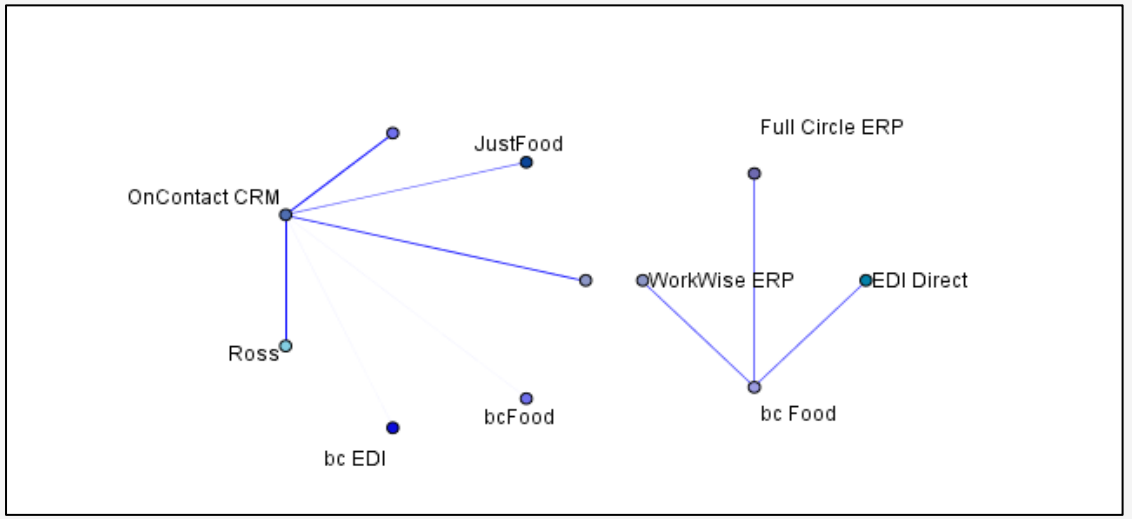
TURF Analysis

"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

REVA UNIVERSITY
Bengaluru, India
Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

Market Basket analysis - Apriori

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



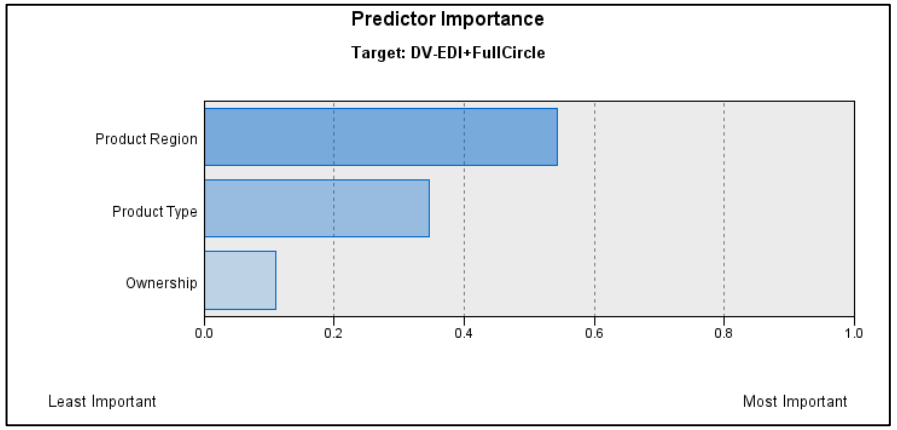
Model Evaluation

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 Precision value of 94%which is, 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%

Decision Tree CHIAD



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

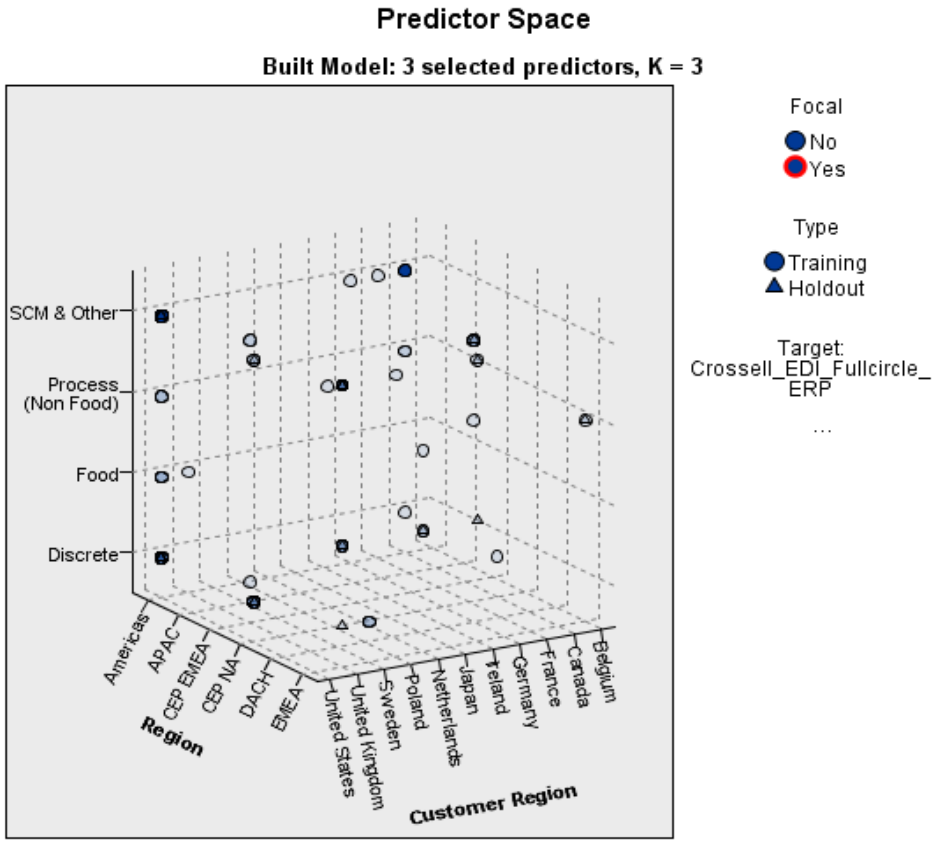
Category	%	n
0.000	100.000	4
1.000	0.000	0
Total	2.000	4

Partnership; Private

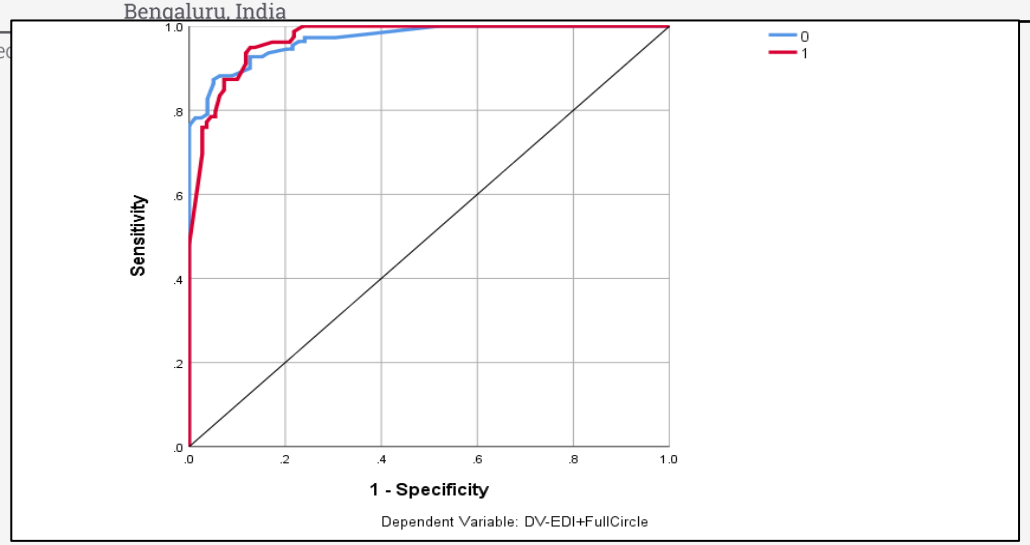
Node 6

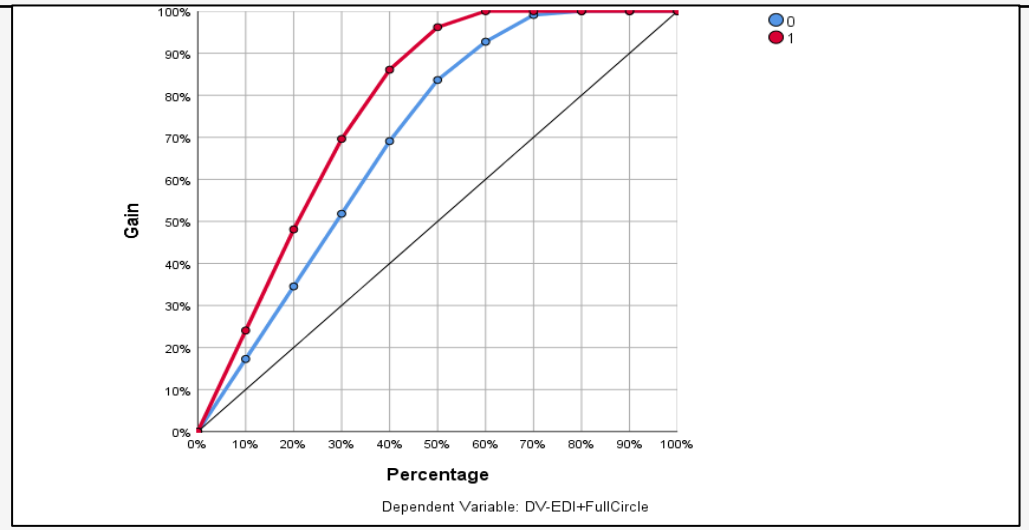
Category	%	n
0.000	27.273	30
1.000	72.727	80
Total	55.000	110

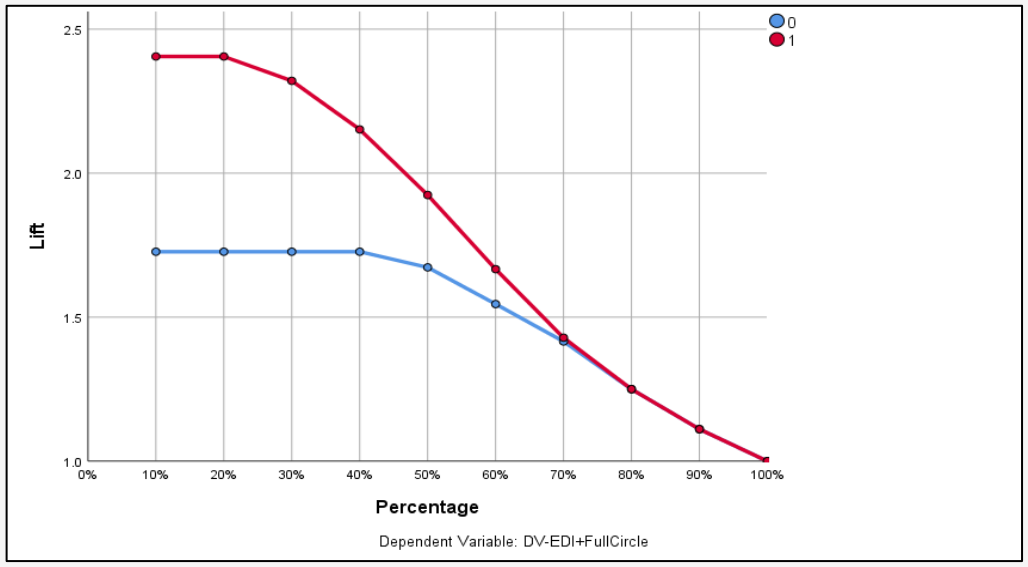
KNN

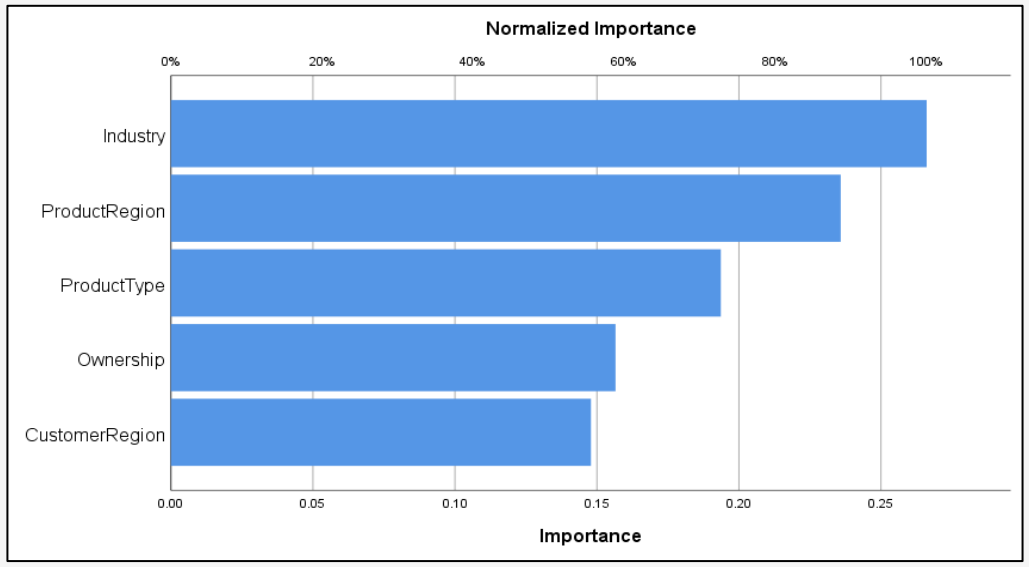

Predictor Space
Built Model: 3 selected predictors, K = 3
SCM & Other
Process_ (Non Food)
Food
Discrete
Region
Americas
APAC
CEP EMEA
CEP NA
DACH
EMEA
Customer Region
United States
United Kingdom
Sweden
Poland
Netherlands
Japan
Ireland
Germany
France
Canada
Belgium
Focal
No
Yes
Type
Training
Holdout
Target: Crossell_EDI_Fullcircle_ ERP
...
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

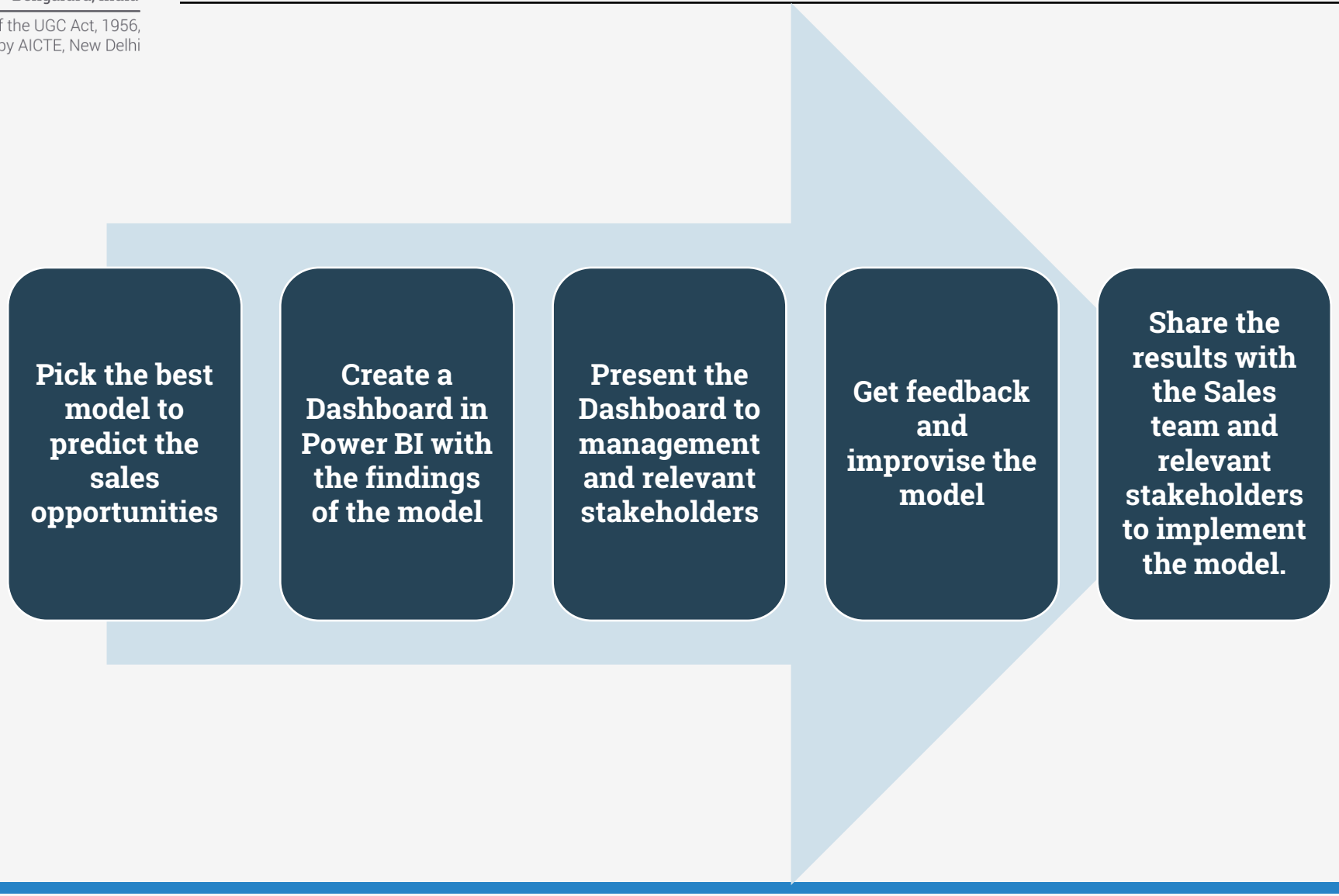








Deploy: Model Deployment Plan



Results and Insights

From the model evaluation though 3 different machine learning approaches were incorporated. The most efficient of the 3 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 the 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

Capitalize

In the last phase of the FLUID mechanism , we have the Capitalize phase, where the findings or the outcome of the study need to be used to increase the revenue and sales for the organization

Conclusion and Future Work

Conclusion:

In this project, different modelling techniques have been tried and evaluated to find out the key drivers responsible for cross-selling certain 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 in organization and enabling the organization to reach its new goal of achieving sales targets and increasing customer base and maintain niche enterprise products.

Scope of future study:

- 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). A 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

Annexure

This project report titled "**Product Affinity Analysis using Machine Learning**" was scanned for similarity detection and the outcome is as given below.

Plagiarism score

Software Used: Turnitin

Date of Report Generation: 26th August 2022

Similarity Index in %: 4%

Total word count: 6751

Plagiarism Report

ORIGINALITY REPORT

4%	3%	0%	3%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

REVA
UNIVERSITY

Bengaluru, India

Established as per the section 2(f) of the UGC Act, 1956,
Approved by AICTE, New Delhi

Thank You!