

# BAYESIAN NETWORK DESIGN FOR A DECISION SUPPORT SYSTEM IN SOUTH ASIAN E-COMMERCE MANAGEMENT

**Ammar Jamshed\***

*\*MSc Student, University of London Goldsmiths*

**\*Corresponding Author:**

## Authors Note

I acknowledge that I have used sources for references of codes and have provided them all in the last section. I confirm that I have provided my own work and not anyone else's.

## Table of Contents

<b>Introduction.....</b>	<b>2</b>
<b>Industry Use-case for Bayesian network used for Decision Support system (DSS). ....</b>	<b>2</b>
<b>Specification of Task use cases and Integrations with other systems.....</b>	<b>3</b>
<b>Specification of Objectives .....</b>	<b>3</b>
<b>Specification of Constraints .....</b>	<b>3</b>
<b>Data Context.....</b>	<b>3</b>
<b>Specification of Components for E-commerce context in Pakistan.....</b>	<b>4</b>
<b>Specification of Bayesian Network. ....</b>	<b>3</b>
<b>Design of Bayesian Network. ....</b>	<b>4</b>
<b>Implementation of Bayesian Network.....</b>	<b>4</b>
<b>Associated Probabilities determined from Pakistan's E-commerce management data analysis.....</b>	<b>6</b>
<b>Selected of Nodes and Links for Bayesian Network.....</b>	<b>6</b>
<b>Querying the Bayesian network. ....</b>	<b>4</b>
<b>Network score.....</b>	<b>7</b>
<b>Description of outputs.....</b>	<b>7</b>
<b>Other usable networks.....</b>	<b>8</b>
<b>Conclusion. ....</b>	<b>5</b>
<b>Potential stakeholders and users.....</b>	<b>8</b>
<b>References .....</b>	<b>5</b>

## Abstract

*E-commerce is a thriving industry in Pakistan which is the a high population density country connected by air travel and road networks to India and China as well as central Asia thus using advanced computing algorithms allow for examination into effective business practices for managerial commerce via smart technology platforms and using Bayesian model for predictability of consumer purchasing based on behavioral variables in Pakistan serve as foundational support in expansion of smart management of E-commerce in Pakistan and other potential markets of Bangladesh and Sri Lanka with similar behavioral trends.*

**Keywords:** *Bayes theorem, E-commerce, South Asia, Analytics.*

## I. INTRODUCTION

The decision Support systems revolve around multiple factors ranging from data structures, and the industrial nature of data collection and analysis to the algorithm for processing the decision support system for industrial nature.

In the case of using Bayesian networks for Decisions Support systems we rely on probabilistic reasoning as well as historical trend linkages to current outcomes based on the type of classification model, we select for developing a Bayesian network in the design of a Decision support system.

Bayesian Network as we know revolves around the use of conditional probability and was derived from the study and work of Thomas Bayes. Today it's used in multiple industries, both dynamic and static industries where we have multiple dependent variables surrounding the business activity and events resulting in outcomes of specific nature.

## II. INDUSTRY USE-CASE FOR BAYESIAN NETWORK USED FOR DECISION SUPPORT SYSTEM (DSS).

South Asian Buying and selling behavior for merchandise and products revolve around complex psychological variables which are dependent and independent variables that are based on the volatility of extrinsic surroundings or stay static regardless of how the surrounding environment transforms.

E-commerce revolves around Understanding buyer behavior about cyber interactivity between the buyer and seller thus we cross many independent and dependent variables within its process ranging from User interface interactivity to Brand and platform reliability within South Asian markets.

The Bayes Theorem formula normally applied to an E-commerce model is:

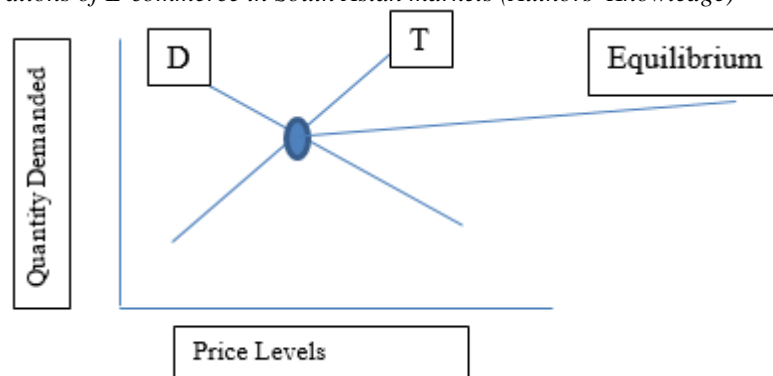
$$P(A/B) = (P(B/A) * P(A)) / P(B)$$

Where:

- $P(A/B)$  is the posterior probability of event A given that B has occurred.
- $P(B/A)$  is the likelihood of event B given that A has occurred.
- $P(A)$  is the prior probability of event A.
- $P(B)$  is the prior probability of event B.

Source (Investopedia)

**Figure 1:** Market Operations of E-commerce in South Asian markets (Authors' Knowledge)



We can examine from the graph that E-commerce buying behavior is based on Intersections between Demand and Trust as people only purchase what they find reliable to do so.

- D is for Demand.
- T is for Trust, which can be company goodwill or other forms of intangible value the E-commerce Brand holds.

In standard classical economics it's assumed that prices and quantity demanded to hold an inverse relationship but in E-commerce economics Trust also intervenes as a third dimension as the higher the trust by consumers over the product the greater the price can be charged over it while the price does impact consumer affordability thus market analysis on consumer income segments have to be examined on the pricing strategy. Thus, where can Demand and Trusted worth in the form of Goodwill or another intangible value intersects is where price determinations start.

## III. SPECIFICATION OF TASK USE CASES AND INTEGRATIONS WITH OTHER SYSTEMS

The Examples of a Decision support system final product for E-commerce management in South Asian business settings can be in the forms of:

- A deployed Bayesian network model on an E-commerce firm's server.
- A generated API of the deployed model.
- A Dashboard of data analytics.
- An application connecting the live data feed to a designed model framework.

#### IV. SPECIFICATION OF OBJECTIVES

Our main aim is to build better predictive analytical solutions for E-commerce in the South Asian markets of Pakistan, Bangladesh, and Sri Lanka to bring benefits of:

- Optimized processes and decision-making.
- Lower costs of product management and analytics.
- Improved business performance.

#### V. SPECIFICATION OF CONSTRAINTS

South Asian markets are not information symmetrical and lack trust in cyber platforms, thus the behavior of Households and markets intersect at the equilibrium of Trust and price. Which is further drilled down to gender and interest when making a purchase decision. This happens due to:

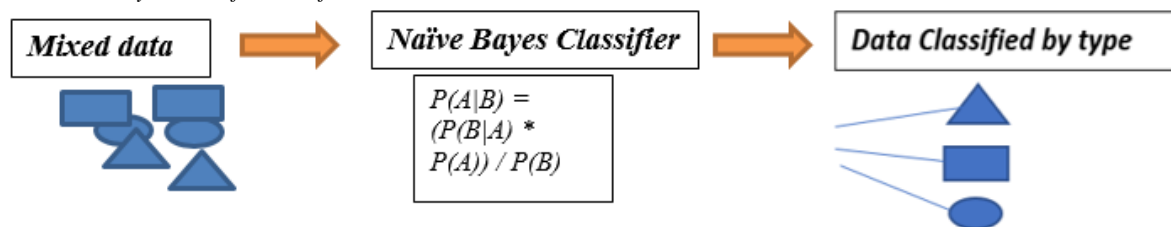
- Lack of technological infrastructure.
- Obsolete data management procedures and storage technologies.
- Lower literacy rates.

#### VI. DATA CONTEXT

The data collected is From the Kaggle repository of Zeeshan Usmani, One of Pakistan's leading data scientists who are working towards revolutionizing data platforms in Pakistan.

The data represents Online Shopping transactions recorded from 2016 to 2018 from multiple merchants selected in a research study in Pakistan and shows Pakistan's potential towards transforming E-commerce and further future spinoffs from this industrial infrastructure catered towards managing it as an industry.

**Figure 1** Naïve Bayes Classifier workflow



Specification of Components for E-commerce context in Pakistan:

Sno	Variable Name	Explanation
1	status	Order status of the transaction
2	price	Price of transaction
3	qty_ordered	Count of Items ordered
4	grand_total	The total amount paid with tax and service charges
5	category_name_1	Category of order
6	payment_method	Type of payment made for order
7	Year	Year of transaction
8	Month	Month of transaction
9	Customer Years (Created after feature engineering)	How long the customer has been with the business
10	Customer ID	Identifier of customers making transactions.

I used Zeeshan Usmani's data to study Pakistan's E-commerce trends and then created random sample data with our set variables as Pakistani E-commerce trends are still far too volatile for Data Accuracy, but we were able to examine how to create a similar network model using it.

#### VII. SPECIFICATION OF BAYESIAN NETWORK.

The Bayesian network to be used within E-commerce would compromise conditionality-induced variables within the given dataset that would require further transformation and encoding before transacting with Bayes Theorem-based machine learning algorithms.

For Pakistan and similar South Asian e-commerce networks, we have conditionalities placed upon trust and reliability of the brand resulting in purchases on platforms thus with Bayes theorem we could predict customer behavior in South Asian markets.

We can view that the Naïve Bayes Classifier clusters data points for classification but it cannot establish historical connections between outcomes as we know  $X^n$  of any dataset will be  $X$  if we use this modeling framework, but we cannot establish within the model lifecycle that  $X^n$  was  $X+2$  before converting into  $X$ .

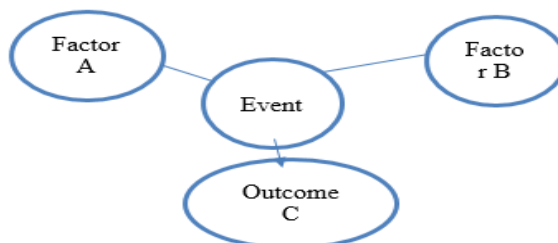
Where:

- $X$  is the present event.
- $X^n$  the number of the same events counted.
- $X+2$  Events before the current events that led to the present event taking place.

### VIII. DESIGN OF BAYESIAN NETWORK.

For Pakistan and similar South Asian e-commerce networks, we have conditionalities placed upon trust and reliability of the brand resulting which are differentiated in terms of gender during purchases on platforms thus with Bayes theorem we could predict customer behavior in South Asian markets.

**Figure 2** Bayesian Network Workflow



We can see that in comparison to Naïve Bayes Classifier, the Bayesian network takes into account pre-conditional events that allow us to examine causality at a much closer range but it takes more computing power and network load of nodes on the servers to run in comparison to Naïve Bayes Classifier. Thus, it makes the Bayesian network the better choice for our modeling approach.

### IX. IMPLEMENTATION OF BAYESIAN NETWORK.

The Bayesian Network was designed and implemented with Python programming using pgmy Library which provided us with a network design algorithm with a reasonable level of computing costs.

We researched E-commerce activity and analyzed via Zeeshan Usmani's datasets and examined the variables we need for our Bayesian network design from it and created sample datasets compromising of those variables with random data and built our model and network using it. This case study must be further researched and inspected by data scientists and E-commerce professionals and my network serves as a strong baseline for its network-building procedure and use cases.

### X. ASSOCIATED PROBABILITIES DETERMINED FROM PAKISTAN'S E-COMMERCE MANAGEMENT DATA ANALYSIS

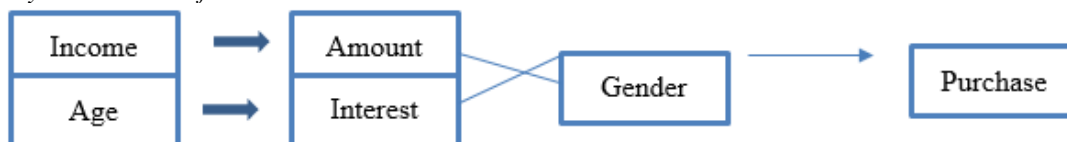
The Logic of our Bayesian Network for E-commerce Activity in South Asia is based on Reliability as a form of Interest and Purchase decisions linked to gender roles in households and income capacity available for purchase.

### XI. SELECTION OF NODES AND LINKS FOR BAYESIAN NETWORK

We examined from Zeeshan Usmani's dataset that the variables sensitive to E-commerce transactions in Pakistan are:

- Gender
- Interest
- Purchase
- Amount
- Age
- Income

**Figure 3** Bayesian Network for South Asian E-commerce



### XII. QUERYING THE BAYESIAN NETWORK.

We query the Bayesian network by eliminating variables to return Discrete factors and looking at how it affects our network reliability in terms of how one variable affects the likelihood of another.

Variable	Average phi(Likelihood)	Query
Purchase	0.25005	Is Purchase important?
Interest	0.2344	Does Gender Affect the network density?
Amount	0.249975	Is amount an important variable for the network?

### XIII. NETWORK SCORE

With Our Bayesian networks performance, we have been able to evaluate the usability of the network on our data and scored its performance with the following estimators:

### XIV. DESCRIPTION OF OUTPUTS

Metric		Value	Explanation
K2Score		-352.630	Learning effectiveness
BDeuScore		-384.104	The fitness of the data to the model
bic.score		-692.593	Model Effectiveness

Due to the selection of random data points for testing the scoring of our model and data do show optimum results but South Asian E-commerce data requires further engineering of fitter data pipelines and transformation to be used with a Bayesian network as random data testing may only be a pilot test for modeling applicability of South Asian E-commerce data with Bayesian networks.

### XV. OTHER USABLE NETWORKS

Normally the Bayes Theorem Algorithm that possibly could be applied would be the Naive Bayes algorithm where we can understand the predictability of the events based on relationships between our features.

### XVI. CONCLUSION.

While we have examined that the cardinality of the model is 4 for each feature. This proves that with Bayesian networks we can predict the effectiveness of South Asian E-commerce giving the right structure of data for probability-linked distributions to apply Bayesian Networks for accurate predictive analytics.

To initiate the development of a Bayesian network and link it to the live feeding point of any E-commerce startup or corporation in South Asia, we need to first transform and wrangle the data variables according to the designed network to run them with trained data science and data engineering teams.

### XVII. POTENTIAL STAKEHOLDERS AND USERS

This will help E-commerce giants like Daraz, Ali Express, Pickaboo, and various other E-commerce startups entering lower developing South Asian Economies engage in more fruitful business management.

### XVIII. REFERENCES

- [1].Clobert, M. (2021). East versus West: Psychology of religion in East Asian cultures. *Current Opinions in Psychology*, 40, 61–66. <https://doi.org/10.1016/j.copsyc.2020.08.021>
- [2].Lee, Y.-H., Hu, P. J.-H., Cheng, T.-H., & Hsieh, Y.-F. (2012). A cost-sensitive technique for positive-example learning supporting content-based product recommendations in B-to-C e-commerce. *Decision Support Systems*, 53(1), 245–256. doi:10.1016/j.dss.2012.01.018
- [3].Ahn, J.-H., & Ezawa, K. J. (1997). Decision support for real-time telemarketing operations through Bayesian network learning. *Decision Support Systems*, 21(1), 17–27. doi:10.1016/S0167-9236(97)00009-2
- [4].Cain, J. D., Jinapala, K., Makin, I. W., Somaratna, P. G., Ariyaratna, B. R., & Perera, L. R. (2003). Participatory decision support for agricultural management. A case study from Sri Lanka. *Agricultural Systems*, 76(2), 457–482. doi:10.1016/S0308-521X(02)00006-9
- [5].Molina, M., Fuentetaja, R., & Garrote, L. (07 2005). Hydrologic Models for Emergency Decision Support Using Bayesian Networks. 88–99. doi:10.1007/11518655\_9
- [6].Alvandi, E. (n.d.). Developing a Web-Based Integrated Bayesian Decision Support System for Rangeland Management in a Semi-Arid Watershed., <http://dx.doi.org/10.2139/ssrn.4174919>
- [7].Mokhtari, K., Ren, J., Roberts, C., & Wang, J. (2012). Decision support framework for risk management on sea ports and terminals using fuzzy set theory and evidential reasoning approach. *Expert Systems with Applications*, 39(5), 5087–5103. doi:10.1016/j.eswa.2011.11.030