# Causal Modeling of E-Commerce Purchasing Intention

**Foad Assareh**
foad.assareh@studio.unibo.it

Master's Degree in Artificial Intelligence, University of Bologna
January 25, 2026

## Abstract

This project models the factors behind online purchasing intention using a Causal Bayesian Network. Using the UCI "Online Shoppers Purchasing Intention Dataset" [1], we developed a causal model to predict user behavior. Despite severe class imbalance (85:15), the model achieved an accuracy of **87%** By utilizing probabilistic inference and a standard decision threshold.

## Introduction

### Domain

We model the transition from "Visitor" to "Buyer" using session metrics (e.g., page views) and context (e.g., weekend).

### Aim

The goal is to build a causal model that predicts purchasing intention and explains *why* purchases fail or succeed. We aim to identify key factors—from behavioral to technical—that influence buying probability.

## Method

We used the `pgmpy` library [2] and the Online Shoppers dataset [1] to implement a Causal Bayesian Network. Methodology included:

- **Preprocessing:** Continuous variables were discretized based on domain logic. Latent variables `UserIntent` and `TechFriction` were derived to model unobserved causal states.
- **Probabilistic Inference:** we utilized probabilistic queries to handle uncertainty. To address data sparsity (rare feature combinations unseen in training), we applied a neutral imputation strategy (filling undefined probabilities with 0.5) during the test phase.
- **Evaluation:** The model was evaluated on a held-out test set using a standard decision threshold of 0.5

## Results

**Performance:** Accuracy = **87%**, Buyer Recall = **81%**. The model captures "Explaining Away": high bounce rates reduce purchase probability even when page views are high.

## Model

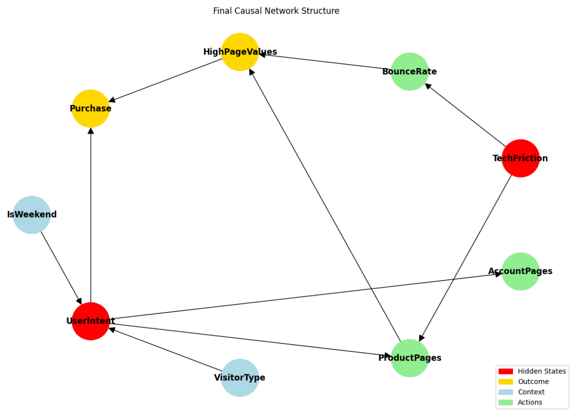

Figure 1: Causal Bayesian Network for predicting purchasing intention.

The network has 9 discrete nodes in four causal layers:

- **Context (Root):** `IsWeekend`, `VisitorType`.
- **State (Hidden):** `UserIntent`, `TechFriction`.
- **Action(Evidence):** `ProductPages`, `AccountPages`, `BounceRate`.
- **Outcome:** `HighPageValues`, `Purchase`.

CPDs were learned from the balanced dataset using Maximum Likelihood Estimation (MLE) [2].

## Analysis

### Experimental Setup

We tested five query scenarios, computing posterior probabilities for `Purchase` given evidence $E$:

1. **Baseline:** $E = \emptyset$.
2. **Strongest Predictor:** $E = \{\texttt{HighPageValues} : Yes\}$.
3. **Explaining Away:** Compare $E_a = \{\texttt{ProductPages} : High\}$ vs $E_b = E_a \cup \{\texttt{BounceRate} : High\}$.
4. **Context Effect:** $E = \{\texttt{IsWeekend} : True\}$.
5. **Ambiguity:** $E = \{\texttt{AccountPages} : Clicked\}$.

### Results

- **Scenario 2 (Strongest Predictor):** The strongest evidence (`HighPageValues:Yes`) raised purchase probability from the baseline of 14.6% to **54.6%**. This confirms that high engagement value is the primary causal driver of purchase.

- **Scenario 3 (Explaining Away):** Initial evidence of interest (`ProductPages:High`) increased probability to **22.2%**. However, observing a high `BounceRate` alongside it caused the probability to drop back to **14.2%** (below baseline). This illustrates the "Explaining Away" effect: the negative signal of bouncing overrides the positive signal of page views.
- **Scenario 4 & 5 (Context):** Contextual factors showed negligible influence; `IsWeekend` resulted in no change from the baseline (**14.6%**), while soft signals like `AccountPages:Clicked` provided a minor lift to **14.9%**.

### Conclusion

- **Robustness:** The Bayesian Network proved naturally robust to class imbalance. It achieved high Recall (0.81) for the minority class using a standard probability threshold, negating the need for oversampling.
- **Causal Insight:** Behavioral metrics (Page Values) significantly outweigh contextual features (Weekend/VisitorType) in predicting purchase intention.

## Links to external resources

- **Code:** GitHub Repository
- **Dataset:** UCI Online Shoppers Intention

## References

[1] Sakar, C.O., et al. (2018). *Online Shoppers Purchasing Intention Dataset.* UCI Machine Learning Repository. `https://archive.ics.uci.edu/ml/datasets/Online+Shoppers+Purchasing+Intention+Dataset`

[2] Ankan, A., & Panda, A. (2015). *pgmpy: Probabilistic Graphical Models using Python.* Documentation: `https://pgmpy.org/`