

---

# Bayesian Prediction of Online Shoppers' Purchasing Intention

---

Liang-Jen Huang<sup>1</sup>

## Abstract

This project uses Bayesian logistic regression to predict whether an online shopping session results in a purchase, based on the Online Shoppers Purchasing Intention Dataset. By modeling both behavioral and contextual variables, the Bayesian framework provides calibrated probabilities and quantifies uncertainty in predictions. The approach offers insights into variation across user groups and visit patterns, supporting a better understanding of online shopping behavior and informing more effective marketing strategies in e-commerce. The code is available at: [github.com/edogawa-liang/bayesian-purchasing-intention-analysis](https://github.com/edogawa-liang/bayesian-purchasing-intention-analysis)

## 1. Introduction

E-commerce companies are constantly trying to understand how customers behave so they can improve marketing, use resources more effectively, and increase sales. A key question is whether a visit to an online store will actually lead to a purchase. Being able to predict this helps businesses focus on the users who are most likely to buy, tailor recommendations, and run more effective campaigns, as shown in previous predictive studies (Moro et al., 2016; Sakar & Kastro, 2018). It also improves the shopping experience while avoiding unnecessary marketing costs.

Many machine learning methods have been used for this type of prediction, but they typically focus on point estimates and accuracy. In contrast, Bayesian modeling provides posterior distributions for each coefficient, allowing uncertainty to be explicitly quantified. It also produces credible intervals for predicted probabilities, offering a clearer picture of confidence in the results. Moreover, hierarchical structures can be incorporated to capture both overall trends and group-specific deviations, resulting in a model that is more interpretable and informative than most conventional machine learning approaches.

In this project, a hierarchical Bayesian logistic regression model with shrinkage priors is applied to analyze the *Online Shoppers Purchasing Intention Dataset* (Sakar & Kastro, 2018), aiming to build a predictive model, identify the key

factors influencing purchase decisions, and quantify the associated uncertainty.

## 2. Data

The analysis is based on the *Online Shoppers Purchasing Intention Dataset* (Sakar & Kastro, 2018), which contains 12,330 sessions collected over a one-year period. Each observation represents the complete browsing activity of a unique user. The target variable is *Revenue*, a binary indicator of whether the session ended in a purchase. In total, the dataset provides 17 features covering both behavioral and contextual aspects of online shopping sessions.

The dataset is well structured, contains no missing values, and includes both numerical and categorical variables. Among all sessions, 84.5% did not result in a purchase, while 15.5% did, leading to an imbalanced target distribution. A summary of the variables is presented in Table 1.

### 2.1. Variable Description

Behavioral variables such as *Administrative*, *Informational*, and *ProductRelated* (and their corresponding durations) describe the number of pages visited and the total time spent in each category. *BounceRates*, *ExitRates*, and *PageValues* are Google Analytics metrics representing user engagement, while *SpecialDay* quantifies the proximity of a visit to a major holiday (e.g., Valentine's Day, Mother's Day). Additional contextual variables include *OperatingSystem*, *Browser*, *Region*, *TrafficType*, *VisitorType*, *Weekend*, and *Month*.

In the subsequent analysis, *Month*, *Region*, and *TrafficType* are treated as grouping factors for the hierarchical Bayesian model to capture both overall trends and group-specific deviations.

### 2.2. Data Preprocessing

The dataset was well structured and required only minimal preprocessing. To remove scale differences among numeric features, z-score standardization was applied so that each variable has a mean of zero and a standard deviation of one, while categorical variables were converted to factors and retained in their original form for analysis.

Table 1. Online Shoppers Purchasing Intention Dataset

Variable	Type
Administrative	Integer
Administrative Duration	Continuous
Informational	Integer
Informational Duration	Continuous
ProductRelated	Integer
ProductRelated Duration	Continuous
BounceRates	Continuous
ExitRates	Continuous
PageValues	Continuous
SpecialDay	Continuous
Month	Categorical
OperatingSystems	Integer
Browser	Integer
Region	Integer
TrafficType	Integer
VisitorType	Categorical
Weekend	Binary
Revenue (Target)	Binary

### 3. Models and Methods

#### 3.1. Bayesian logistic regression

A Bayesian logistic regression model was applied to predict the target variable *Revenue*. Let  $y_i$  denote the purchase outcome for session  $i$ , where  $y_i = 1$  if the session ended in a purchase and  $y_i = 0$  otherwise:

$$y_i \sim \text{Bernoulli}(\pi_i), \quad \text{logit}(\pi_i) = \alpha + \mathbf{x}_i^\top \boldsymbol{\beta},$$

where  $\pi_i$  is the purchase probability,  $\alpha$  is the intercept, and  $\boldsymbol{\beta}$  are coefficients for predictors  $\mathbf{x}_i$ .

The model was implemented in Stan through the RStan interface and estimated using the No-U-Turn Sampler (NUTS). Four Markov chains were run, each with 2,000 iterations including 1,000 warm-up steps.

#### POOLED MODEL

In the pooled model, all sessions share the same intercept and coefficients, assuming no group-specific variation. Weakly informative priors were assigned to regularize the estimation:

$$y_i \sim \text{Bernoulli}(\pi_i), \quad \text{logit}(\pi_i) = \alpha + \mathbf{x}_i^\top \boldsymbol{\beta}, \\ \alpha \sim \mathcal{N}(0, 5^2), \quad \beta_j \sim \mathcal{N}(0, 2^2).$$

This specification assumes that all data points are exchangeable and that any potential group-level differences are absorbed into the shared parameters.

#### HIERARCHICAL MODEL

To account for heterogeneity among groups (e.g., *Month*, *Region*, or *TrafficType*), a hierarchical model was constructed with varying intercepts. Each observation  $i$  belongs to a group  $g[i]$ , and the model allows the intercept to vary by group:

$$y_i \sim \text{Bernoulli}(\pi_i), \\ \text{logit}(\pi_i) = \alpha + \mathbf{x}_i^\top \boldsymbol{\beta} + \alpha_{g[i]}, \\ \alpha_g \sim \mathcal{N}(0, \sigma_{\text{group}}), \\ \sigma_{\text{group}} \sim \text{Exponential}(1).$$

The group-level intercepts  $\alpha_g$  introduce partial pooling, borrowing strength across groups to prevent overfitting and to stabilize estimates for smaller categories. This structure captures systematic variation between groups while retaining overall population-level regularization.

#### SHRINKAGE PRIORS (BAYESIAN LASSO)

To improve generalization and perform variable selection, a shrinkage prior was applied to the regression coefficients. Specifically, a Laplace (double-exponential) prior was used, corresponding to the Bayesian analogue of the LASSO (Park & Casella, 2008):

$$\beta_j \sim \text{Laplace}(0, \lambda^{-1}), \quad \lambda \sim \text{Exponential}(1).$$

The global shrinkage parameter  $\lambda$  controls the strength of regularization: larger  $\lambda$  values imply stronger shrinkage (coefficients are pulled closer to zero), while smaller  $\lambda$  values indicate weaker shrinkage, allowing the model to remain more flexible. The Laplace prior's heavier tails, compared to a Gaussian prior, enable it to shrink small coefficients toward zero while preserving larger effects.

In this study, the Bayesian LASSO prior was integrated into the hierarchical structure, allowing both group-level partial pooling and coefficient shrinkage.

#### 3.2. Model Assessment and Comparison

To evaluate and compare the predictive performance of the candidate models, the *Pareto-smoothed importance sampling leave-one-out cross-validation* (PSIS-LOO) was employed, as WAIC yielded similar results but PSIS-LOO is more accurate and provides better diagnostic measures.

**Leave-One-Out cross-validation (LOO).** The expected log predictive density ( $\text{elpd}_{\text{loo}}$ ) was used as the main criterion for model comparison, with higher values indicating better predictive accuracy. It measures how well the model

predicts unseen data and is computed as

$$\widehat{\text{elpd}}_{\text{loo}} = \sum_{i=1}^n \log \left( \frac{1}{S} \sum_{s=1}^S p(y_i | \theta^{(s)})^{w_i^{(s)}} \right),$$

where  $w_i^{(s)}$  are Pareto-smoothed importance weights for observation  $i$  and posterior draw  $s$ . The effective number of parameters  $p_{\text{loo}}$  reflects the variability of the pointwise log-likelihoods, and the overall information criterion is defined as

$$\text{LOOIC} = -2 \widehat{\text{elpd}}_{\text{loo}}.$$

A smaller LOOIC (or equivalently, a larger  $\widehat{\text{elpd}}_{\text{loo}}$ ) indicates better out-of-sample predictive performance. The Pareto- $k$  diagnostic was also examined to assess the reliability of the approximation, with values  $k < 0.7$  generally considered acceptable.

## 4. Results

### 4.1. Model comparison

Six Bayesian logistic regression models were fitted to the on-line shoppers dataset. The first model was a pooled logistic regression without hierarchical structure. Three single-level hierarchical models were then estimated, each including varying intercepts for one grouping variable: *Month*, *Region*, or *TrafficType*. Next, a two-level hierarchical model with varying intercepts for both *Region* and *TrafficType* was fitted. Finally, a hierarchical model with the same two grouping variables but with shrinkage priors (Bayesian LASSO) on the regression coefficients was estimated.

As shown in Tables 2, the PSIS-LOO results provide a comparison of the models' out-of-sample predictive performance. All models exhibited stable Pareto- $k$  diagnostics ( $k < 0.7$ ), except for the hierarchical model with Month grouping (*H\_M*), which showed two slightly higher values. This suggests that the PSIS-LOO estimates are reliable and not notably affected by influential observations.

Table 2. Model comparison using PSIS-LOO.

Model	elpd_loo	p_loo	LOOIC	Pareto- $k$ status
P (Pooled)	-3603.4	61.1	7206.9	all < 0.7
H_M	-3604.2	60.7	7208.3	2 in (0.7, 1.0]
H_T	-3603.0	59.2	7206.1	all < 0.7
H_R	-3600.5	55.8	7201.0	all < 0.7
H_R+T	-3597.7	38.7	7195.4	all < 0.7
H_R+T+L	<b>-3593.4</b>	<b>37.6</b>	<b>7186.8</b>	all < 0.7

Note. P = Pooled; H\_M = Hierarchical (Month); H\_T = Hierarchical (TrafficType); H\_R = Hierarchical (Region); H\_R+T = Hierarchical (Region + Traffic); H\_R+T+L = Hierarchical (Region + Traffic, LASSO).

From Tables 2 and 3, adding hierarchical structures based on *Region* or *TrafficType* improved predictive performance ( $\widehat{\text{elpd}}_{\text{loo}} = -3603.0$  and  $-3600.5$ ) compared with the pooled model ( $\widehat{\text{elpd}}_{\text{loo}} = -3603.4$ ). This indicates that accounting for group-level differences helps capture heterogeneity in user behavior. Since using *Month* as a grouping factor did not lead to improvement, both *Region* and *TrafficType* were next included as hierarchical levels. This two-level structure further improved predictive accuracy ( $\widehat{\text{elpd}}_{\text{loo}} = -3597.7$ ), suggesting that multiple grouping variables better capture cross-group variability. Finally, adding shrinkage priors (*H\_R+T+L*) produced the best overall performance, yielding the highest  $\widehat{\text{elpd}}_{\text{loo}}$  ( $-3593.4$ ) and the lowest  $p_{\text{loo}}$  (37.6), indicating better generalization with reduced model complexity. The shrinkage also made parameter estimates more stable and helped mitigate potential overfitting.

Table 3. Differences in expected log predictive density ( $\widehat{\text{elpd}}_{\text{loo}}$ ) relative to the best model (H\_R+T+L).

Model	elpd_diff (SE)
H_R+T+L	0.0 (0.0)
H_R+T	-4.3 (1.8)
H_R	-7.1 (4.7)
H_T	-9.6 (4.9)
P	-10.0 (5.2)
H_M	-10.8 (5.2)

Note. P = Pooled; H\_M = Hierarchical (Month); H\_T = Hierarchical (TrafficType); H\_R = Hierarchical (Region); H\_R+T = Hierarchical (Region + Traffic); H\_R+T+L = Hierarchical (Region + Traffic, LASSO).

### 4.2. Posterior Estimates

The final model is a hierarchical Bayesian logistic regression with two grouping variables, *Region* and *TrafficType*, and shrinkage priors (Bayesian LASSO) on the regression coefficients. The model is specified as:

$$\begin{aligned} y_i &\sim \text{Bernoulli}(p_i), \\ \text{logit}(p_i) &= \alpha + \mathbf{x}_i^\top \boldsymbol{\beta} + \alpha_{\text{region}[i]}^{(1)} + \alpha_{\text{traffic}[i]}^{(2)}, \\ \alpha &\sim \mathcal{N}(0, 5), \\ \lambda &\sim \text{Exponential}(1), \\ \beta_j &\sim \text{Laplace}(0, \lambda^{-1}), \quad j = 1, \dots, K, \\ \sigma_{\text{region}}, \sigma_{\text{traffic}} &\sim \text{Exponential}(1), \\ \alpha_{\text{region}[r]}^{(1)} &\sim \mathcal{N}(0, \sigma_{\text{region}}), \quad r = 1, \dots, J_1, \\ \alpha_{\text{traffic}[t]}^{(2)} &\sim \mathcal{N}(0, \sigma_{\text{traffic}}), \quad t = 1, \dots, J_2. \end{aligned}$$

The posterior means were  $\hat{\alpha} = -1.86$ ,  $\hat{\lambda} = 1.16$  (corresponding to a Laplace scale of  $\lambda^{-1} = 0.86$ ),  $\hat{\sigma}_{\text{region}} = 0.30$ , and  $\hat{\sigma}_{\text{traffic}} = 0.06$ .

Key posterior estimates are reported in Table 4, and the full list of parameter summaries is available in Appendix A. According to the convergence diagnostics, the model showed satisfactory convergence. The  $\hat{R}$  values and the trace plots displayed good mixing and stability across chains, as detailed in Appendix A and Appendix B.

Table 4. Posterior summaries of key parameters discussed in the text.

Parameter	Mean	SD	95% CI
$\alpha$ (Global intercept)	-1.86	0.21	[-2.26, -1.45]
$\beta_7$ (BounceRates)	-1.5	1.6	[-5.3, 0.8]
$\beta_8$ (ExitRates)	-12.7	1.9	[-16.5, -8.9]
$\beta_9$ (PageValues)	0.0827	0.0024	[0.0780, 0.0880]
$\beta_{11}$ (Month Dec)	-0.60	0.17	[-0.94, -0.27]
$\beta_{12}$ (Month Feb)	-1.4	0.6	[-2.6, -0.4]
$\beta_{17}$ (Month Nov)	0.54	0.15	[0.26, 0.84]
$\sigma_{\text{region}}$	0.06	0.05	[0.00, 0.18]
$\sigma_{\text{traffic}}$	0.30	0.10	[0.15, 0.53]
$\lambda$ (Shrinkage)	0.86	0.20	[0.55, 1.30]

#### 4.3. Interpretation

The posterior summaries in Appendix A provide detailed insights into the hierarchical Bayesian logistic regression with *Region* and *TrafficType* group-level effects under Bayesian LASSO shrinkage. The global intercept ( $\alpha = -1.86$ ) corresponds to an average baseline purchase probability of approximately 13.5%,

$$p = \frac{e^{-1.86}}{1 + e^{-1.86}} \approx 0.135.$$

reflecting the overall tendency toward non-purchase outcomes in the dataset.

Among the continuous predictors, *BounceRates* ( $\beta_7 = -1.50$ ) and *ExitRates* ( $\beta_8 = -12.70$ ) exhibit strong negative associations with purchase likelihood, indicating that users who leave pages quickly or exit frequently are considerably less likely to complete a transaction. In contrast, *PageValues* ( $\beta_9 = 0.083$ ) shows a clear positive effect, suggesting that higher page value scores are associated with a greater probability of purchase.

Regarding the time-related variables, visits in December (*Month Dec*) and February (*Month Feb*) are associated with lower purchase probabilities, while visits in November (*Month Nov*) have a positive effect, indicating that purchasing behavior varies slightly across months.

The estimated group-level standard deviations indicate moderate variability across groups:  $\sigma_{\text{region}} = 0.06$  and  $\sigma_{\text{traffic}} = 0.30$ . This implies that differences between regions are

relatively small, while differences across traffic sources contribute more to the model's uncertainty.

The global shrinkage parameter ( $\lambda^{-1} = 0.86$ ) reflects moderate regularization, allowing the Bayesian LASSO to down-weight less informative predictors while preserving key behavioral signals (e.g., *BounceRates*, *ExitRates*) and preventing overfitting.

## 5. Conclusions

This study applied hierarchical Bayesian logistic regression to model online shoppers' purchasing intention using both behavioral and contextual features. The results show that incorporating hierarchical structures for *Region* and *TrafficType*, together with Bayesian LASSO shrinkage, provided the best predictive performance among the tested models. By combining partial pooling and shrinkage, the model was able to capture cross-group variation while avoiding overfitting.

The analysis also revealed several important predictors of purchase behavior, such as *BounceRates*, *ExitRates*, and *ProductRelated Duration*, which reflect users' engagement and intent during browsing sessions. Moreover, explicitly modeling group-level factors like *Region* and *TrafficType* helped account for contextual differences in purchasing patterns across user segments.

Overall, the hierarchical Bayesian framework not only provides full posterior distributions and credible intervals for all parameters, but also offers interpretable insights into online shopping behavior, supporting more data-driven and personalized strategies in e-commerce.

**Future work.** Future work could explore advanced Bayesian machine learning methods such as BART and compare them with Bayesian statistical learning approaches. It would also be interesting to examine whether Bayesian techniques can outperform traditional machine learning in predictive performance.

## References

- Moro, S., Cortez, P., and Rita, P. A data-driven approach to predict the success of bank telemarketing. *Decision Support Systems*, 62:22–31, 2016. doi: 10.1016/j.dss.2014.03.001.
- Park, T. and Casella, G. The bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686, 2008. doi: 10.1198/016214508000000337.
- Sakar, C. and Kastro, Y. Online Shoppers Purchasing Intention Dataset. UCI Machine Learning Repository, 2018. DOI: <https://doi.org/10.24432/C5F88Q>.

## Appendix A. Posterior Parameter Estimates

Table 5: Posterior estimates from the hierarchical Bayesian logistic regression with *Region* and *TrafficType* group-level effects under Bayesian LASSO shrinkage.

Parameter	Mean	MCSE	SD	2.5%	97.5%	$\hat{R}$
Global intercept	-1.86	0.0069	0.21	-2.26	-1.45	1.00
Administrative	0.007	0.0002	0.011	-0.014	0.027	1.00
Administrative Duration	-0.00012	0.000003	0.00019	-0.00051	0.00025	1.00
Informational	0.027	0.0004	0.027	-0.026	0.080	1.00
Informational Duration	0.00004	0.000003	0.00023	-0.00041	0.00048	1.00
Product Related	0.0022	0.00002	0.0012	-0.0002	0.0044	1.00
Product Related Duration	0.00006	0.0000004	0.00003	0.00001	0.00012	1.00
Bounce Rates	-1.5	0.036	1.6	-5.3	0.8	1.00
Exit Rates	-12.7	0.030	1.9	-16.5	-8.9	1.00
Page Values	0.0827	0.00002	0.0024	0.078	0.088	1.00
Special Day	-0.14	0.0036	0.22	-0.6	0.26	1.00
Month (Dec)	-0.60	0.0055	0.17	-0.94	-0.27	1.00
Month (Feb)	-1.4	0.0115	0.6	-2.6	-0.4	1.00
Month (Jul)	0.14	0.0052	0.19	-0.24	0.52	1.00
Month (Jun)	-0.23	0.0054	0.24	-0.73	0.22	1.00
Month (Mar)	-0.47	0.0055	0.17	-0.79	-0.14	1.00
Month (May)	-0.47	0.0054	0.16	-0.78	-0.16	1.00
Month (Nov)	0.54	0.0052	0.15	0.26	0.84	1.00
Month (Oct)	0.03	0.0054	0.18	-0.32	0.38	1.00
Month (Sep)	0.06	0.0053	0.19	-0.30	0.43	1.00
Operating Systems	-0.067	0.0007	0.038	-0.143	0.006	1.00
Browser	0.038	0.0003	0.019	0.001	0.076	1.00
Visitor Type (Other)	-0.38	0.0078	0.45	-1.38	0.42	1.00
Visitor Type (Returning)	-0.26	0.0016	0.09	-0.44	-0.09	1.00
Weekend (TRUE)	0.08	0.0012	0.07	-0.05	0.22	1.00
Region 1	0.00	0.0011	0.04	-0.08	0.10	1.00
Region 2	0.04	0.0024	0.06	-0.05	0.20	1.01
Region 3	0.00	0.0011	0.05	-0.10	0.10	1.00
Region 4	-0.01	0.0014	0.05	-0.14	0.09	1.00
Region 5	-0.02	0.0019	0.07	-0.20	0.08	1.00
Region 6	0.01	0.0014	0.06	-0.10	0.14	1.00
Region 7	0.00	0.0013	0.05	-0.11	0.12	1.00
Region 8	0.01	0.0014	0.06	-0.11	0.15	1.00
Region 9	-0.03	0.0025	0.07	-0.22	0.07	1.01
Traffic Type 1	-0.07	0.0033	0.12	-0.31	0.17	1.00
Traffic Type 2	0.10	0.0032	0.11	-0.11	0.33	1.00
Traffic Type 3	-0.28	0.0033	0.13	-0.54	-0.04	1.00
Traffic Type 4	-0.01	0.0034	0.14	-0.28	0.26	1.00
Traffic Type 5	0.12	0.0036	0.17	-0.22	0.47	1.00
Traffic Type 6	-0.11	0.0033	0.17	-0.46	0.22	1.00
Traffic Type 7	0.08	0.0040	0.25	-0.41	0.58	1.00
Traffic Type 8	0.37	0.0042	0.17	0.06	0.72	1.00
Traffic Type 9	-0.03	0.0045	0.27	-0.60	0.51	1.00
Traffic Type 10	0.21	0.0037	0.16	-0.08	0.54	1.00
Traffic Type 11	0.18	0.0036	0.18	-0.16	0.55	1.00

(continued on next page)

(continued from previous page)

Parameter	Mean	MCSE	SD	2.5%	97.5%	$\hat{R}$
Traffic Type 12	0.00	0.0049	0.32	-0.65	0.65	1.00
Traffic Type 13	-0.46	0.0040	0.18	-0.84	-0.13	1.00
Traffic Type 14	-0.04	0.0054	0.30	-0.68	0.56	1.00
Traffic Type 15	-0.15	0.0052	0.30	-0.81	0.40	1.00
Traffic Type 16	0.07	0.0052	0.32	-0.52	0.77	1.00
Traffic Type 17	-0.01	0.0047	0.32	-0.67	0.61	1.00
Traffic Type 18	-0.04	0.0052	0.32	-0.73	0.58	1.00
Traffic Type 19	-0.07	0.0045	0.30	-0.69	0.52	1.00
Traffic Type 20	0.19	0.0043	0.20	-0.20	0.60	1.00
$\sigma_{\text{region}}$	0.06	0.0030	0.05	0.00	0.18	1.02
$\sigma_{\text{traffic}}$	0.30	0.0031	0.10	0.15	0.53	1.00
$\lambda$ (Global shrinkage)	0.86	0.0035	0.20	0.55	1.30	1.00

## Appendix B. Convergence Diagnostics using Traceplots

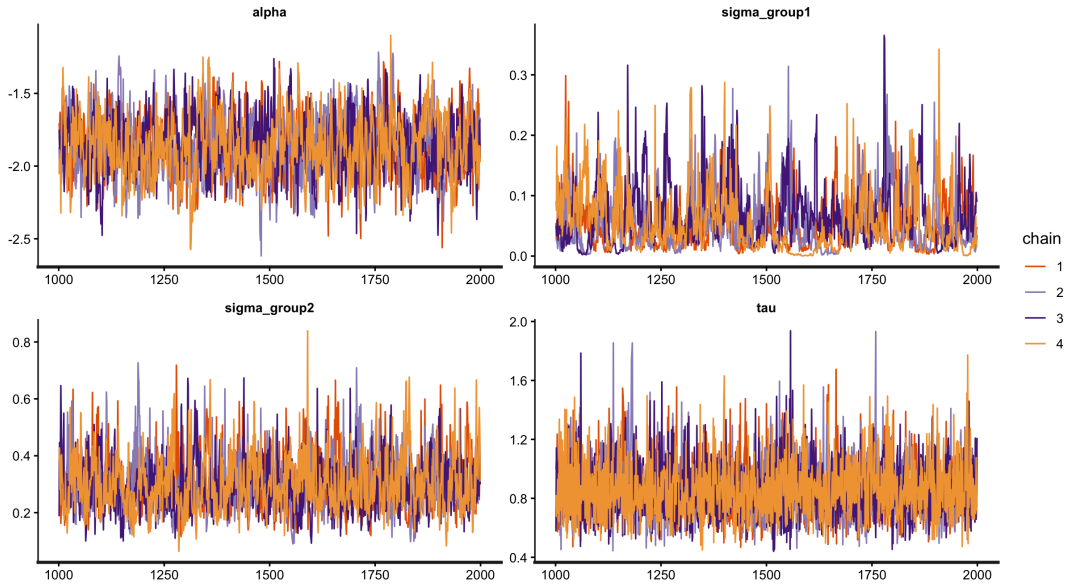


Figure 1. Traceplots of key parameters in the hierarchical Bayesian logistic regression model with *Region* and *TrafficType* as grouping variables and LASSO shrinkage. Four MCMC chains are shown for each parameter after warm-up.

The traceplots in Figure 1 show good mixing and stable fluctuations for the key parameters ( $\alpha$ ,  $\sigma_{\text{region}}$ ,  $\sigma_{\text{traffic}}$ , and  $\lambda$ ), with  $\hat{R} \approx 1.00$  for most, indicating satisfactory convergence.  $\sigma_{\text{region}}$  shows a slightly higher value ( $\hat{R} = 1.02$ ), which merits attention but remains within an acceptable range.

## Appendix C. Stan Code

```
data {
  int<lower=0> N;
  int<lower=0> K;
  matrix[N, K] X;
  int<lower=0, upper=1> y[N];
}
parameters {
```

```

real alpha;
vector[K] beta;
}
model {
  alpha ~ normal(0, 5);
  beta ~ normal(0, 2);
  y ~ bernoulli_logit(alpha + X * beta);
}
generated quantities {
  vector[N] log_lik;
  for (n in 1:N)
    log_lik[n] = bernoulli_logit_lpmf(y[n] | alpha + dot_product(X[n], beta));
}

```

Listing 1. Stan code for pooled Bayesian logistic regression

```

data {
  int<lower=0> N;
  int<lower=0> K;
  int<lower=1> J;
  int<lower=1,upper=J> group_id[N];
  matrix[N, K] X;
  int<lower=0,upper=1> y[N];
}
parameters {
  real alpha;
  vector[K] beta;
  vector[J] alpha_group;
  real<lower=0> sigma_group;
}
model {
  alpha ~ normal(0, 5);
  beta ~ normal(0, 2);
  sigma_group ~ exponential(1);
  alpha_group ~ normal(0, sigma_group);
  y ~ bernoulli_logit(alpha + X * beta + alpha_group[group_id]);
}
generated quantities {
  vector[N] log_lik;
  for (n in 1:N) {
    real eta = alpha + dot_product(X[n], beta) + alpha_group[group_id[n]];
    log_lik[n] = bernoulli_logit_lpmf(y[n] | eta);
  }
}

```

Listing 2. Stan code for hierarchical Bayesian logistic regression (one grouping variable)

```

data {
  int<lower=0> N;
  int<lower=0> K;
  int<lower=1> J1;
  int<lower=1> J2;
  int<lower=1,upper=J1> group1_id[N];
  int<lower=1,upper=J2> group2_id[N];
  matrix[N, K] X;
  int<lower=0,upper=1> y[N];
}
parameters {
  real alpha;
  vector[K] beta;
  vector[J1] alpha_group1;
  vector[J2] alpha_group2;
  real<lower=0> sigma_group1;
  real<lower=0> sigma_group2;
}

```

```

}
model {
  alpha ~ normal(0, 5);
  beta ~ normal(0, 2);
  sigma_group1 ~ exponential(1);
  sigma_group2 ~ exponential(1);
  alpha_group1 ~ normal(0, sigma_group1);
  alpha_group2 ~ normal(0, sigma_group2);
  y ~ bernoulli_logit(alpha + X * beta +
                     alpha_group1[group1_id] +
                     alpha_group2[group2_id]);
}
generated quantities {
  vector[N] log_lik;
  for (n in 1:N) {
    real eta = alpha + dot_product(X[n], beta) +
               alpha_group1[group1_id[n]] +
               alpha_group2[group2_id[n]];
    log_lik[n] = bernoulli_logit_lpmf(y[n] | eta);
  }
}

```

Listing 3. Stan code for hierarchical Bayesian logistic regression with two grouping variables (Region and TrafficType)

```

data {
  int<lower=0> N;
  int<lower=0> K;
  int<lower=1> J1;
  int<lower=1> J2;
  int<lower=1,upper=J1> group1_id[N];
  int<lower=1,upper=J2> group2_id[N];
  matrix[N, K] X;
  int<lower=0,upper=1> y[N];
}
parameters {
  real alpha;
  vector[K] beta;
  vector[J1] alpha_group1;
  vector[J2] alpha_group2;
  real<lower=0> sigma_group1;
  real<lower=0> sigma_group2;
  real<lower=0> tau; // global shrinkage parameter (Bayesian LASSO)
}
model {
  // Priors
  alpha ~ normal(0, 5);
  tau ~ exponential(1); // global shrinkage
  beta ~ double_exponential(0, tau); // Laplace prior (LASSO)

  sigma_group1 ~ exponential(1);
  sigma_group2 ~ exponential(1);
  alpha_group1 ~ normal(0, sigma_group1);
  alpha_group2 ~ normal(0, sigma_group2);

  // Likelihood
  for (n in 1:N) {
    real eta = alpha + dot_product(X[n], beta)
               + alpha_group1[group1_id[n]]
               + alpha_group2[group2_id[n]];
    y[n] ~ bernoulli_logit(eta);
  }
}
generated quantities {
  vector[N] log_lik;
}

```

```
for (n in 1:N) {  
  real eta = alpha + dot_product(X[n], beta)  
    + alpha_group1[group1_id[n]]  
    + alpha_group2[group2_id[n]];  
  log_lik[n] = bernoulli_logit_lpmf(y[n] | eta);  
}  
}
```

*Listing 4.* Stan code for hierarchical Bayesian logistic regression with two grouping variables and Bayesian LASSO shrinkage priors