Predictive Modeling: Click Behaviour on Advertisement

Group 3:

1. Miecel Alicia Angel J - 2702327601
2. Natasha Kayla Cahyadi - 2702235891
3. Sherly Vaneza - 2702222163

Dataset Used : Dataset Advertisement - Clicked on Ad
Code in Rmd : OneDrive - Code & Dataset

Introduction

Advertisement effectiveness is a key area of interest in marketing and business analytics. Advertisements play a crucial role in driving **consumer behavior, increasing brand awareness, and boosting sales**. The goal of this project is to develop a probabilistic model using Bayesian methods to predict whether a user will click on an advertisement. By analyzing the relationship between user attributes (**such as age, daily time spent on site, area income, and daily internet usage**) and the likelihood of clicking on ads, the project aims to provide insights into user behavior and improve targeted advertising strategies.

For our **Mini Project**, The dataset contains **1000 rows and 10 columns**, providing data about user behavior on an advertising platform. Below is the summary of the structure and content:

1. **Daily Time Spent on Site:** Time spent on the website (minutes).
2. **Age:** Consumer's age (years).
3. **Area Income:** Average income in the consumer's region.
4. **Daily Internet Usage:** Minutes spent online daily.
5. **Ad Topic Line:** Advertisement headline.
6. **City/Country:** Consumer's location.
7. **Male:** Consumer's gender (1 = Male, 0 = Female).
8. **Timestamp:** Time of ad interaction.

Target (Y): *Clicked on Ad* (1 = Yes, 0 = No).

MODELS

1. **Logistic Regression - Informative Prior**

   Logistic regression is an appropriate choice for predicting the *Clicked on Ad* variable because it is specifically designed to model **binary outcomes**. For instance, a study by **Facebook researchers** combined decision trees with logistic regression to predict ad clicks, highlighting the importance of selecting appropriate features and models for accurate predictions. Additionally, incorporating informative priors in Bayesian logistic regression enhances predictions. Google researchers found that regularized logistic regression is well-suited for large-scale ad click prediction.

2. **Probit Regression - Uninformative Prior**

Probit regression is a statistical method used to predict **binary outcomes**, such as whether an advertisement will be clicked, by linking predictor variables to the probability of an outcome through the **cumulative distribution function (CDF)** of the standard normal distribution. While uninformative priors allow the data to drive the analysis, they can sometimes lead to unintuitive results. Incorporating weakly informative priors can help guide the model without overwhelming the data. For instance, Gelman et al. (2008) recommend using a Student-t prior distribution with a scale parameter to provide gentle regularization in regression models, which can improve parameter estimation and model convergence.

Algorithm

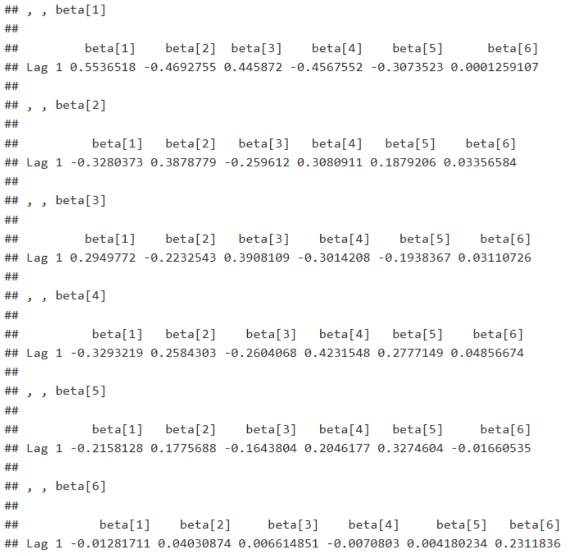
1. Logistic Regression

```
model_str <- "
model {
  for (i in 1:N) {
   Y[i] ~ dbern(pi[i])
    logit(pi[i]) <- beta[1] + inprod(X[i, ], beta[2:(p + 1)])

   loglike[i] <- Y[i] * log(max(pi[i], 1e-10)) + (1 - Y[i]) * log(max(1 -
pi[i], 1e-10))
  }

  for (j in 1:(p + 1)) {
    beta[j] ~ dnorm(mu[j], tau[j])
  }
}"
```

In our **Logistic Regression**, we uses:

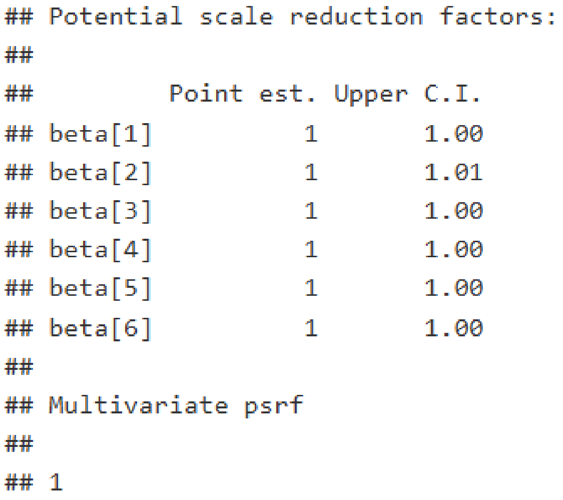
- **Likelihood:** Bernoulli Distribution, predicting probabilities of *Clicked on Ad* bounded between 0 and 1.
- **Priors:** Normal Distribution with informative priors, incorporates prior knowledge from historical data to improve model accuracy.

Autocorrelation



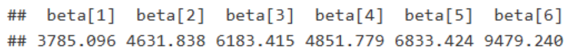
```
## , , beta[1]
##
##          beta[1]    beta[2]  beta[3]    beta[4]    beta[5]      beta[6]
## Lag 1 0.5536518 -0.4692755 0.445872 -0.4567552 -0.3073523 0.0001259107
##
## , , beta[2]
##
##           beta[1]   beta[2]   beta[3]   beta[4]   beta[5]    beta[6]
## Lag 1 -0.3280373 0.3878779 -0.259612 0.3080911 0.1879206 0.03356584
##
## , , beta[3]
##
##          beta[1]    beta[2]   beta[3]    beta[4]    beta[5]    beta[6]
## Lag 1 0.2949772 -0.2232543 0.3908109 -0.3014208 -0.1938367 0.03110726
##
## , , beta[4]
##
##           beta[1]   beta[2]    beta[3]   beta[4]   beta[5]    beta[6]
## Lag 1 -0.3293219 0.2584303 -0.2604068 0.4231548 0.2777149 0.04856674
##
## , , beta[5]
##
##           beta[1]   beta[2]    beta[3]   beta[4]   beta[5]     beta[6]
## Lag 1 -0.2158128 0.1775688 -0.1643804 0.2046177 0.3274604 -0.01660535
##
## , , beta[6]
##
##            beta[1]    beta[2]     beta[3]    beta[4]     beta[5]   beta[6]
## Lag 1 -0.01281711 0.04030874 0.006614851 -0.0070803 0.004180234 0.2311836
```

Gelman Diag



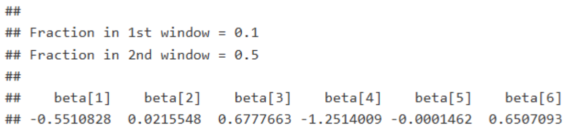
```
## Potential scale reduction factors:
##
##         Point est. Upper C.I.
## beta[1]          1       1.00
## beta[2]          1       1.01
## beta[3]          1       1.00
## beta[4]          1       1.00
## beta[5]          1       1.00
## beta[6]          1       1.00
##
## Multivariate psrf
##
## 1
```

ESS



```
##   beta[1]  beta[2]  beta[3]  beta[4]  beta[5]  beta[6]
## 3785.096 4631.838 6183.415 4851.779 6833.424 9479.240
```

Geweke Diag 2



```
##
## Fraction in 1st window = 0.1
## Fraction in 2nd window = 0.5
##
##    beta[1]    beta[2]    beta[3]    beta[4]    beta[5]    beta[6]
## -0.5510828  0.0215548  0.6777663 -1.2514009 -0.0001462  0.6507093
```

Image 1. Logistic Regression Convergence Diagnostic Result

The **Logistic Regression Model** successfully converged, revealing that higher area income increases the likelihood of clicking ads, while older age, being male, increased internet usage, and more time on-site reduce it, offering reliable insights for optimizing advertising strategies.

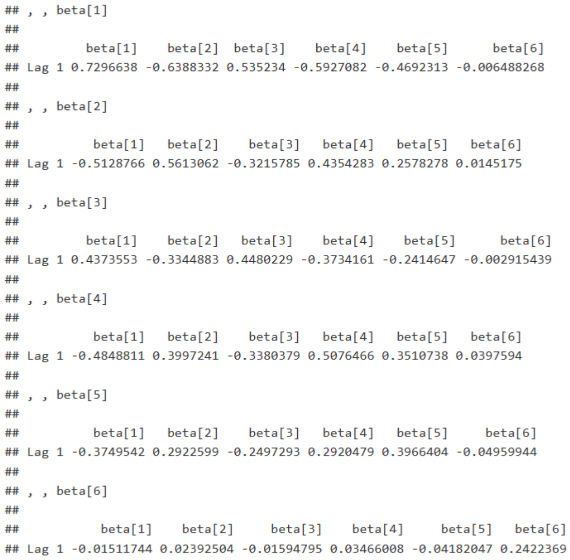
2. Probit Regression

```
modelstr2 <- textConnection("model {
  for (i in 1:N) {
    Y[i] ~ dbern(pi[i])

    pi[i] <- phi(beta[1] + inprod(X[i, ], beta[2:(p + 1)]))

    loglike[i] <- Y[i] * log(max(pi[i], 1e-10)) + (1 - Y[i]) * log(max(1 -
pi[i], 1e-10))
  }

  for (j in 1:(p + 1)) {
    beta[j] ~ dnorm(0, 0.01)
  }
}")
```

In our **Probit Regression**, we uses:

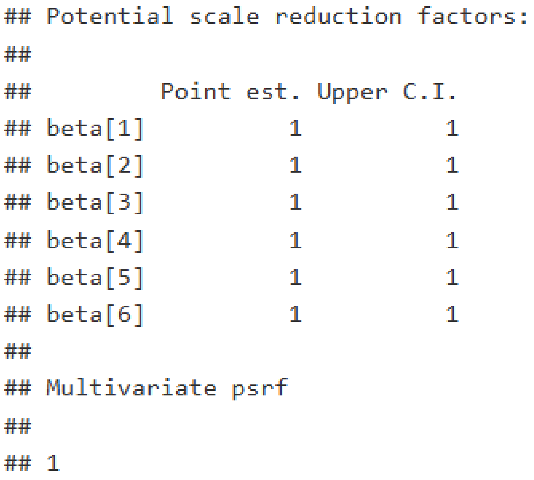
- **Likelihood:** Bernoulli distribution, predicting probabilities of *Clicked on Ad* with a logit function.
- **Priors:** Normal Distribution with uninformative priors.

Autocorrelation



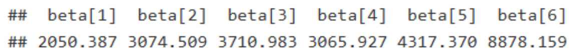
```
## , , beta[1]
##
##          beta[1]    beta[2]  beta[3]    beta[4]    beta[5]      beta[6]
## Lag 1 0.7296638 -0.6388332 0.535234 -0.5927082 -0.4692313 -0.006488268
##
## , , beta[2]
##
##           beta[1]   beta[2]    beta[3]   beta[4]   beta[5]   beta[6]
## Lag 1 -0.5128766 0.5613062 -0.3215785 0.4354283 0.2578278 0.0145175
##
## , , beta[3]
##
##          beta[1]    beta[2]   beta[3]    beta[4]    beta[5]      beta[6]
## Lag 1 0.4373553 -0.3344883 0.4480229 -0.3734161 -0.2414647 -0.002915439
##
## , , beta[4]
##
##           beta[1]   beta[2]    beta[3]   beta[4]   beta[5]   beta[6]
## Lag 1 -0.4848811 0.3997241 -0.3380379 0.5076466 0.3510738 0.0397594
##
## , , beta[5]
##
##           beta[1]   beta[2]    beta[3]   beta[4]   beta[5]     beta[6]
## Lag 1 -0.3749542 0.2922599 -0.2497293 0.2920479 0.3966404 -0.04959944
##
## , , beta[6]
##
##            beta[1]    beta[2]     beta[3]    beta[4]     beta[5]   beta[6]
## Lag 1 -0.01511744 0.02392504 -0.01594795 0.03466008 -0.04182047 0.2422369
```

Gelman Diag



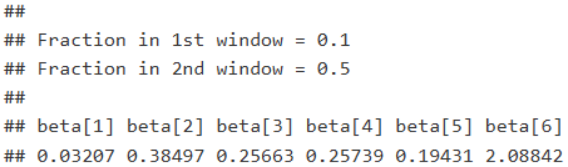
```
## Potential scale reduction factors:
##
##         Point est. Upper C.I.
## beta[1]          1          1
## beta[2]          1          1
## beta[3]          1          1
## beta[4]          1          1
## beta[5]          1          1
## beta[6]          1          1
##
## Multivariate psrf
##
## 1
```

ESS



```
##   beta[1]  beta[2]  beta[3]  beta[4]  beta[5]  beta[6] 
## 2050.387 3074.509 3710.983 3065.927 4317.370 8878.159 
```

Geweke Diag 2



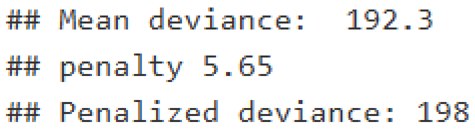
```
## 
## Fraction in 1st window = 0.1
## Fraction in 2nd window = 0.5 
## 
## beta[1] beta[2] beta[3] beta[4] beta[5] beta[6] 
## 0.03207 0.38497 0.25663 0.25739 0.19431 2.08842 
```

Image 2. Probit Regression Convergence Diagnostic Result

The **Probit Regression Model** performs well, as shown by Gelman and Geweke diagnostics, ESS, and autocorrelation results. Most parameters show **good or moderate convergence** with sufficient ESS and acceptable autocorrelation, ensuring reliable estimates. While beta[1] (intercept) has slightly higher autocorrelation and lower ESS, beta[6] (daily time on site) demonstrates efficient sampling. Overall, the model is robust and offers reliable insights into ad-click behavior, making it suitable for targeted advertising.

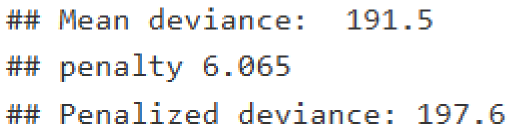
Results

DIC Model 1



```
## Mean deviance:  192.3 
## penalty 5.65 
## Penalized deviance: 198 
```

DIC Model 2



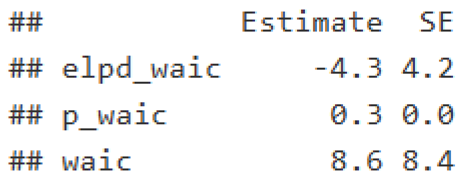
```
## Mean deviance:  191.5 
## penalty 6.065 
## Penalized deviance: 197.6 
```

Image 3. Deviance Information Criterion Result

For Model 1, the mean deviance is 192.3, reflecting its fit to the data, with a penalty for complexity of 5.65. This results in a penalized deviance (DIC) of 198. On the other hand, Model 2 has a mean deviance of 191.5, indicating a slightly better fit to the data. However, it has a higher complexity penalty of 6.065, leading to a DIC of 197.6.

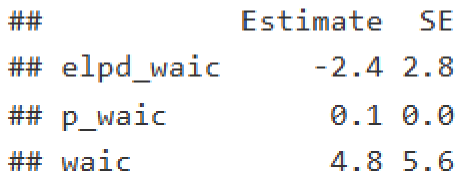
When comparing the two models, **Model 2 demonstrates a better overall performance** with a lower DIC value (197.6) compared to Model 1 (198). While Model 2 is slightly more complex, it's better fit to the data justifies the additional complexity. Therefore, for the advertising dataset, **Model 2 is the preferred choice as it offers a more optimal balance between data fit and model complexity.**

WAIC Model 1



```
##           Estimate  SE
## elpd_waic     -4.3 4.2
## p_waic         0.3 0.0
## waic           8.6 8.4
```

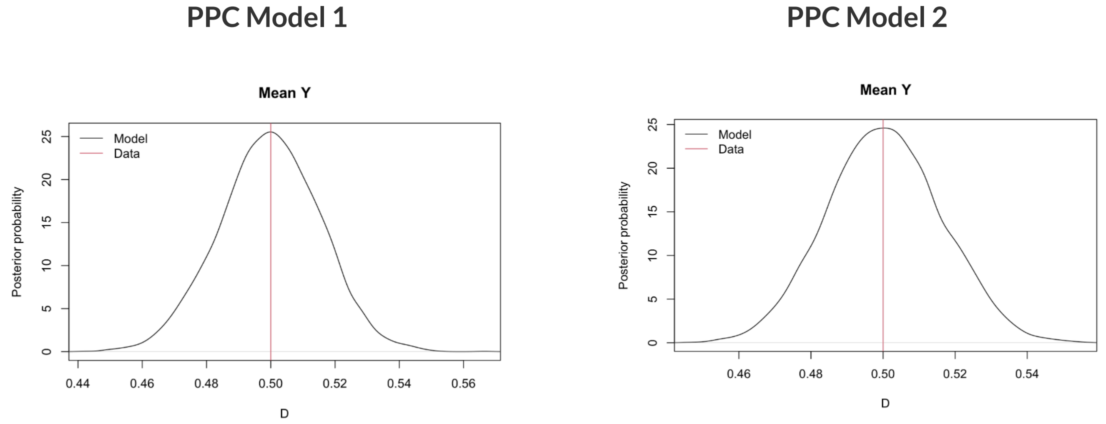
WAIC Model 2



```
##           Estimate  SE
## elpd_waic     -2.4 2.8
## p_waic         0.1 0.0
## waic           4.8 5.6
```

Image 4. Watanabe-Akaike Information Criterion Result

The second model has a lower WAIC (4.8) compared to the first model (8.6), indicating better predictive performance. Although both models have relatively high standard errors (8.4 for the first and 5.6 for the second), the notable difference in WAIC suggests that **the second model is more likely to generalize well to new data.**



PPC Model 1

PPC Model 2

Image 5. Posterior Predictive Checks Result

Both models closely match the observed mean (0.5), indicating that they fit the data well. However, Model 2 (0.505) is slightly closer to the observed mean than Model 1 (0.4982), suggesting a marginally better alignment with the observed data. **Although the difference is minimal, Model 2 demonstrates a more accurate fit and is therefore the preferred choice for the advertisement dataset.**

Conclusion

Based on the combined evaluation using DIC and WAIC, **Model 2 is the preferred choice for the advertising dataset** as it strikes a better balance between data fit, complexity, and predictive performance. Furthermore, considering the PPC results, **Model 2 demonstrates a more accurate fit for this dataset**, solidifying it as the best choice.

Additionally, from the summary and plot of Model 2, the variables impact the target as follows:

- **Age**: Older individuals are less likely to click on ads.
- **Area Income**: Higher income areas increase the likelihood of clicking on ads.
- **Daily Internet Usage**: More internet usage decreases the likelihood of clicking on ads.
- **Gender**: Males are less likely to click on ads than females.
- **Daily Time on Site**: Spending more time on the site slightly decreases the likelihood of clicking on ads.

References

He, X., Pan, J., Jin, O., Xu, T., Liu, B., Xu, T., Shi, Y., Atallah, A., Herbrich, R., Bowers, S., & Candela, J. Q. (2014). Practical lessons from predicting clicks on ads at facebook. *Proceedings of the Eighth International Workshop on Data Mining for Online Advertising*, 1–9. https://doi.org/10.1145/2648584.2648589

McMahan, H. B., Holt, G., Sculley, D., Young, M., Ebner, D., Grady, J., Nie, L., Phillips, T., Davydov, E., Golovin, D., Chikkerur, S., Liu, D., Wattenberg, M., Hrafnkelsson, A. M., Boulos, T., & Kubica, J. (2013). Ad click prediction. *Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. https://doi.org/10.1145/2487575.2488200