Examining the Factors Influencing the Transition of Filipino Actors to Politics Using Logistic Regression Analysis

Baybayon, Darlyn Antoinette B.

Mayol, Jose Raphael J.

Bachelor of Science in Applied Mathematics

APM1205: Linear Models

Mr. Alvarina, Jeffrey

May 16, 2025

Chapter 1

Introduction

1.1 Background of the Study

With the intention of promoting inclusivity, the 1987 Philippine Constitution does not impose restrictions on prior political experience or educational background for individuals running for public office. This allows almost anyone to vie for a government seat, even celebrities who may lack relevant education or experience. Historically, celebrities have played a significant role in Philippine politics as campaign endorsers for politicians, leveraging their mass appeal to boost candidates' visibility and popularity. Eventually, celebrities themselves ran for public office, capitalizing on the fame they accrued as film and TV stars. Rogelio dela Rosa, a 1950s matinee idol, was the first film celebrity to become a Senator in 1958. Later in 1998, action star Joseph Estrada rose to presidency. Currently, in the 19th Congress, five out of twenty-four Senate seats are held by media personalities.

1.2 Statement of the Problem

Although the phenomenon of actors transitioning from entertainment to politics is not distinct to this country, its persistence across all levels of government likely is. This presents a unique opportunity to examine the factors that drive celebrities to pursue and win political positions.

1.3 Objectives of the Study

This study aims to identify the factors significantly associated with actors becoming elected politicians and examine the strength and direction of their influence using logistic regression modeling.

1.4 Hypothesis

Null Hypothesis: The selected predictors (years active, acting credits, awards, family in politics, and educational level) do not significantly affect the likelihood of an actor to run for public office and win.

1.5 Significance of the Study

Existing quantitative research on celebrities in politics examine voters' profiles rather than celebrities' (Atun & David, 2015; David & San Pascual, 2016). The findings of this study will provide valuable insights into this topic by analyzing the characteristics of celebrities who may or may not seek a government position.

1.6 Scope and Limitations

The analysis will include their total number of acting credits in television and film, years of activity in the entertainment industry, number of awards received, educational attainment, and number of immediate family members in politics.

Data collection is done manually through compilation from various sources which may lead to potential discrepancies. The population of actor-turned-politicians is inherently limited, resulting in a relatively small sample size.

Chapter II

Review of Related Literature

Political dynasties are persistent in the Philippines with about 80% of the Congress and over 50% of all elected local government officials hailing from political families (Mendoza et al., 2022). A study on voter preference in senatorial candidates found that those who vote for senators from political dynasties are more likely to vote for celebrities (David & Legara, 2015). This highlights the advantage of name recall for successful electoral bids. As a result, some celebrities with both personal fame and prominent political family names gain more leverage.

Celebrities appeal to the public through emotion or the "construction of a public personality" (McKernan, 2011). Vitug (2004) suggests that with the public's disappointment and waning trust in traditional politicians, celebrities may be viewed as refreshing alternatives.

Educational attainment is on the top of mind among Filipino voters, claiming its reflection on candidate's ideals and political prowess as a crucial factor in assessing their suitability for public office (Abiera et al., 2022). Hence, despite most celebrities being unable to finish college, those who aspire to run often opt to do so, leveraging their popularity and credentials to create a strong image that is both relatable and qualified.

Chapter III

Methodology

3.1 The Dataset

A purposive sampling approach was used to gather as many actors-turned-politicians as possible, as this group is relatively limited. Purposive random sampling was applied to select a comparable number of actors who had not entered politics. To ensure similar levels of public exposure, we imposed the following selection criteria:

- The actor must have at least 10 acting credits.
- The actor must have been active in the entertainment industry at any point between 2005 and 2025.

The following information was collected for each actor:

- years_active: Numerical; Number of years active as an actor until 2025
- imdb_credits: Numerical; Number of TV and film acting credits
- wins: Numerical; Total number of wins for Best Actor and Best Supporting Actor categories from prominent award-giving bodies (Filipino Academy of Movie Arts and Sciences (FAMAS), Gawad Urian, and Metro Manila Film Festival (MMFF))
- family: Numerical; Number of immediate family members in politics
- education: Categorical; Levels: 0 – Not finished basic education, 1 – Basic Education, 2 – Bachelor's Degree, 3 – Master's Degree, 4 – Doctorate
- politics: Categorical, target variable; 1 – Elected politician at any point until May 2022 elections, 0 – Not elected.

Data were obtained from and manually cross-checked across the following online sources: IMDb (for acting credits and awards), Wikipedia (for career duration, political involvement, family, and educational background), Geni.com (for identifying immediate family members) and News websites such as Inquirer, Philippine Star, PEP, etc. (for validation of other information). The dataset had 124 observations in total.

3.2 Data Preprocessing and Transformation

Data preprocessing began by addressing missing values in the *education* variable. The ten observations with missing *education* data were imputed using the median value of the variable. Following imputation, the *education* variable was converted into a factor to prepare for analysis. Diagnostics were then conducted on the regression model base dataset. As six outliers appeared that were too extreme even for transformation, they were removed without any resulting issues. The resulting dataset of 118 observations was used to train the new logistic regression model.

3.3 Model Building

A logistic regression model was constructed in R with *years_active*, *imdb_credits*, *family*, *education*, & *wins* as predictors, and *politics* as the target variable. This method was chosen since the target variable is a binary outcome. The relationship between the predictors and the outcome is represented by the log odds (logit) of the outcome's probability, where each coefficient shows how much the log odds of the outcome change with a one-unit increase in the corresponding predictor (Siavoshi, 2024).

3.4 Assumption Checks

To assess whether logistic regression is a model fitted for the intended purpose and the dataset, its six key assumptions were checked as follows:

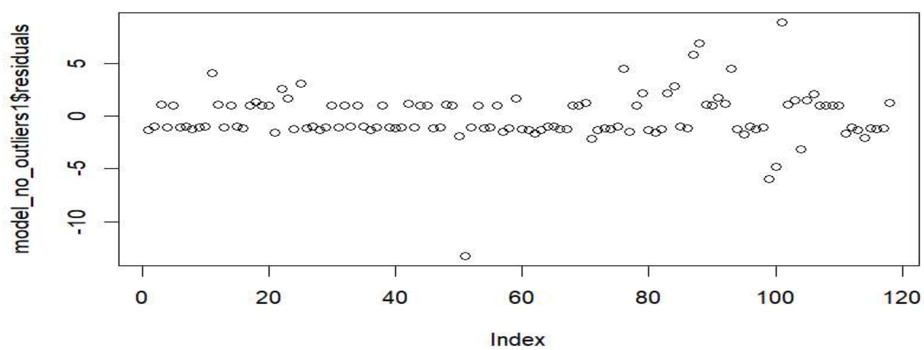
3.4.1 Binary Response Variable

The response variable *politics* takes on only two values — 0 or 1. Thus, the assumption of a binary response variable is met.

3.4.2 Independent Observations

Figure 1

Plot of residuals against time



Since the points only exhibit a random pattern, then the assumption is met.

3.4.3 No Multicollinearity

To detect multicollinearity, the variance inflation factors (VIFs) of the predictor variables were taken.

Table 1

VIFs of predictor variables

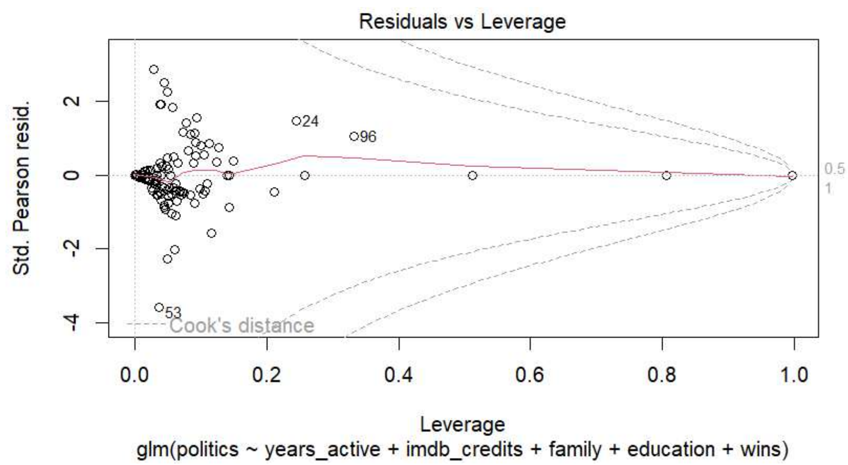
	GVIF	Df	GVIF^(1/(2*Df))
years_active	1.974741	1	1.405255
imdb_credits	2.041041	1	1.428650
family	1.134300	1	1.065035
education	1.063636	4	1.007741
wins	1.392589	1	1.180080

As shown in Table 1, *family*, *education*, and *wins* show almost no multicollinearity (VIF ≈ 1) while *years_active* and *imdb_credits* have moderate values (1<VIF<5). These are all acceptable numbers, and hence, the assumption is met.

3.4.4 No Extreme Outliers

Figure 2

Residuals vs. Leverage plot



The plot indicates that there are no extreme outliers or highly influential points. While three observations exhibit moderate deviation from the fitted values, they do not surpass Cook’s distance thresholds. Hence, the model satisfies the assumption of no extreme outliers.

3.4.5 Linear Relationship Between Explanatory Variables and Logit of the Response Variable

Table 2

Box-Tidwell test results

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -18.15376 2389.03386  -0.008 0.993937
years_active  -0.32654    0.52031  -0.628 0.530282
log_yrs        0.09554    0.11914   0.802 0.422627
imdb_credits  -0.09949    0.10673  -0.932 0.351229
log_imdb       0.01485    0.01966   0.755 0.449969
family         2.15723    0.58296   3.700 0.000215 ***
education1    19.22092 2389.03199   0.008 0.993581
education2    19.70713 2389.03194   0.008 0.993418
education3    35.84972 3249.80693   0.011 0.991198
education4    40.14880 3251.22302   0.012 0.990147
wins          -0.98671    0.39683  -2.486 0.012902 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Since none of the log-transformed variables are significant, then the assumption is met.

3.4.6 Large Enough Sample Size

While the dataset does not meet conventional sample size expectations for logistic regression, this limitation is a consequence of studying a naturally small and bounded population. Despite this, all other key assumptions have been satisfied. Thus, analysis of the model remains valid and meaningful within the scope of the population being studied.

3.5 Model Summary

The logistic regression model summary was generated using the built-in *summary()* function in R. Coefficient estimates, standard errors, and p-values would be used to explain the model's behavior. Additionally, the AIC would be compared to that of the unprocessed model, which is 127.7.

3.6 Model Evaluation

Model performance was evaluated using a confusion matrix and related classification metrics, including accuracy, sensitivity, specificity, positive predictive value (PPV), and the Kappa statistic. The model's predictive reliability was further assessed using McNemar's test and balanced accuracy.

Chapter IV

Results and Discussion

4.1 Interpretations

Table 3

Fitted logistic regression model summary

```
Call:
glm(formula = politics ~ years_active + imdb_credits + family +
    education + wins, family = binomial, data = df_no_outliers1)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  -21.68727 2425.75982  -0.009 0.992867
years_active   0.09013    0.03646   2.472 0.013441 *
imdb_credits  -0.01791    0.01018  -1.760 0.078337 .
family         2.12217    0.55867   3.799 0.000146 ***
education1    19.25220 2425.75968   0.008 0.993668
education2    19.67544 2425.75964   0.008 0.993528
education3    35.60365 3295.98625   0.011 0.991381
education4    39.67097 3198.57518   0.012 0.990104
wins          -0.97875    0.38555  -2.539 0.011130 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 160.176  on 117  degrees of freedom
Residual deviance:  76.005  on 109  degrees of freedom
AIC: 94.005

Number of Fisher Scoring iterations: 17
```

The results in Table 3 indicate that *years_active* (B= 0.09013, p=0.013), *wins* (B=-0.97875, p=0.011), and *family* (B=2.12217, p=0.000146) are statistically significant predictors. Specifically, the *family* variable is a very significant predictor compared to the rest. Interestingly, the *wins* variable is negatively associated with the *politics* response variable. The predictors *imdb_credits* (p=0.078) and *education* (p=0.99) did not reach statistical significance, with the former being marginally significant and the latter having extremely high standard errors and coefficients. Additionally, the AIC for this model is 94.005, substantially lower than the 127.7 AIC of the original unprocessed version.

Table 4

Fitted logistic regression model confusion matrix

```
Confusion Matrix and Statistics

          Reference
Prediction  0  1
         0 62 10
         1  7 39

               Accuracy : 0.8559
                 95% CI : (0.7794, 0.9138)
    No Information Rate : 0.5847
    P-Value [Acc > NIR] : 1.713e-10

                  Kappa : 0.7007

 Mcnemar's Test P-Value : 0.6276

            Sensitivity : 0.8986
            Specificity : 0.7959
         Pos Pred Value : 0.8611
         Neg Pred Value : 0.8478
             Prevalence : 0.5847
         Detection Rate : 0.5254
   Detection Prevalence : 0.6102
      Balanced Accuracy : 0.8472

       'Positive' Class : 0
```

Table 4 shows the confusion matrix of the logistic regression model. The model achieved an accuracy of 85.59% (95% CI: 77.94%-91.4%), significantly outperforming the No Information Rate of 58.5% ($p<0.001$). The balanced accuracy was 84.72%, indicating good performance across both classes. The model showed high sensitivity (89.96%), correctly identifying most celebrities who are not involved or did not win in politics (class 0). It also achieved a positive predictive value (PPV) of 86.1%, meaning that about 86% of those predicted as class 0 were correctly classified. Substantial agreement between predicted and actual outcomes is supported by a Kappa of 0.7007, and McNemar's test ($p=0.6276$) indicates no difference in disagreement between classes.

Chapter V

Conclusion and Recommendations

5.1 Conclusion

The findings of the study reveal that among celebrity aspirants, those with family ties to political dynasties are by far the most likely to successfully make the transition from entertainment to elected office. This variable stood out as the most powerful predictor in the model, with a p-value well below 0.001. In a country where dynastic politics has long been celebrated, celebrity status becomes significantly more potent when coupled with an inherited political network, the dual advantage of fame and lineage. Name recall is essential in Philippine elections, and when a household name is paired with a recognizable political surname, the path to candidacy is almost surefire.

Years active in the entertainment industry also emerged as a significant predictor, suggesting that sustained media presence may serve as an informal form of political capital. The longer a celebrity remains visible and relevant, the more embedded they become in the public psyche, increasing their viability when making the switch into politics. Surprisingly, winning awards was negatively associated with a celebrity's odds to run for and win a position in public office. One plausible interpretation is that highly awarded celebrities may be more invested in their artistic careers and less likely to divert their attention to politics. Alternatively, voters may relate more to celebrities who are perceived as accessible rather than exceptional.

In contrast, IMDB credits and educational attainment were not statistically significant. The former approached only marginal significance, suggesting that breadth of work may contribute weakly to political viability. Education was strikingly non-significant, with inflated standard errors and coefficient estimates. Although education did not emerge as a statistically significant predictor in the model, it clearly influences predicted probabilities. Celebrities with postgraduate degrees (class 3 or 4) show near-certain probabilities of securing political seats, while those without basic education (class 0) show near-zero probabilities. There were simply not enough observations to fill in those educational extremes. This reflects the broader structural reality of postgraduate degrees being exceedingly rare among celebrities unless they have political aspirations, and modern personalities without basic education being few and far between. The lack in statistical power of the variable is a symptom of this data sparsity, not evidence that education lacks relevance as a predictor.

Despite these challenges, the model proved highly effective at distinguishing politically “winnable” celebrities from those who are not. Its strong and balanced performance across all classes suggests that the underlying predictors, particularly the significant ones, capture meaningful patterns in political entry. Taken together, the results emphasize that celebrity-to-politician trajectories in the Philippines are shaped more by social and cultural capital rather than formal accolades. In a political landscape molded by populist appeal, fame and lineage remain the most valuable currencies for turning visibility into electoral success.

5.2 Recommendations

The structure of the celebrity population constrains representation at key ends of the spectrum especially in education, contributing to data sparsity and inflated values in this predictor. Moreover, this analysis did not account for other potentially influential factors such as media sentiment or political messaging. Future research could benefit from larger and more comprehensive datasets to reduce sparsity-related limitations, improve model stability, and potentially open better insights into education as a predictor. Finally, drawing from social media and public perception data or even face-to-face survey data could offer a better understanding of the real dynamics that shape successful celebrity-to-politician trajectories.

References

Abiera, M. A. C., Angeles, J. L., Arboso, J., Brosas, J. A. M., Logrosa, R.L., Moreto, D. J. H., Murillo, M. C. C., Pinangay, V. F. O., & Villamor, C. J. A. (2022). Voting preferences in the national election: A Case among first-time voters. *International Journal of Innovative Science and Research Technology, 7*(6), 1234-1243. https://doi.org/10.38124/ijisrt

Atun, J. M. L., & David, C. C. (2015). Celebrity politics: correlates of voting for celebrities in Philippine presidential elections. Social Science Diliman: *A Philippine Journal of Society and Change, 11*(2), 1-23.

David, C. C., & Legara, E. F. T. (2015). How voters combine candidates on the ballot: the case of the Philippine senatorial election. *International Journal of Public Opinion Research, 29*(1), 70-94. https://doi.org/10.1093/ijpor/edv041

David, C. C., & San Pascual, M. R. S. (2016). Predicting vote choice for celebrity and political dynasty candidates in Philippine national elections. *Philippine Political Science Journal, 37*(2), 82-93. https://doi.org/10.1080/01154451.2016.1198076

McKernan, B. (2011). Politics and Celebrity: A Sociological understanding. *Sociology Compass*, *5*(3), 190–202. https://doi.org/10.1111/j.1751-9020.2011.00359.x

Siavoshi, M. (2024, October 28). *Step-by-Step Guide to Logistic Regression in R*. Statology. https://www.statology.org/step-by-step-guide-to-logistic-regression-in-r/

Vitug, M. D. (2004, February 20). Celebrity politics : Star power holds perils for the Philippines. *The New York Times*. https://www.nytimes.com/2004/02/20/opinion/celebrity-politics-star-power-holds-perils-for-the-philippines.html

Appendices

Figures

Figure 1

Plot of residuals against time

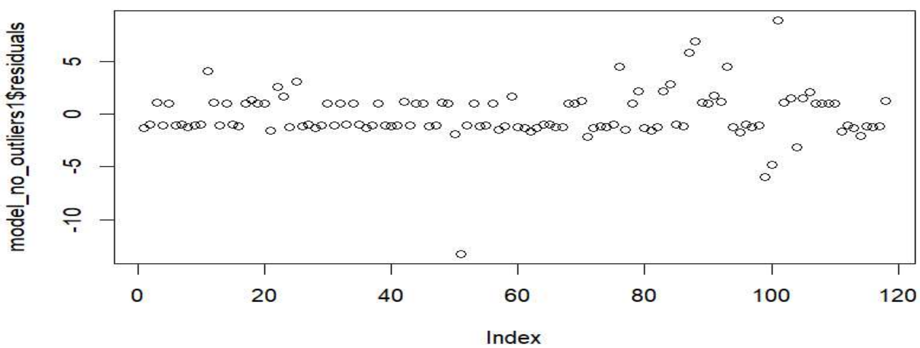
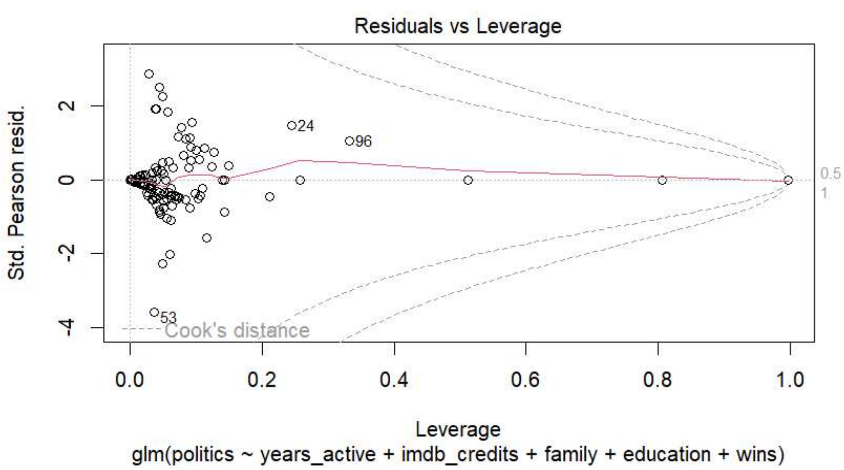


Figure 2

Residuals vs. Leverage plot



Tables

Table 1

VIFs of predictor variables

```
                GVIF Df GVIF^(1/(2*Df))
years_active 1.974741  1        1.405255
imdb_credits 2.041041  1        1.428650
family       1.134300  1        1.065035
education    1.063636  4        1.007741
wins         1.392589  1        1.180080
```

Table 2

Box-Tidwell test results

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)    
(Intercept)  -18.15376 2389.03386  -0.008 0.993937    
years_active  -0.32654    0.52031  -0.628 0.530282    
log_yrs        0.09554    0.11914   0.802 0.422627    
imdb_credits  -0.09949    0.10673  -0.932 0.351229    
log_imdb       0.01485    0.01966   0.755 0.449969    
family         2.15723    0.58296   3.700 0.000215 ***
education1    19.22092 2389.03199   0.008 0.993581    
education2    19.70713 2389.03194   0.008 0.993418    
education3    35.84972 3249.80693   0.011 0.991198    
education4    40.14880 3251.22302   0.012 0.990147    
wins          -0.98671    0.39683  -2.486 0.012902 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Table 3

Fitted logistic regression model summary

```
Call:
glm(formula = politics ~ years_active + imdb_credits + family + 
    education + wins, family = binomial, data = df_no_outliers1)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)    
(Intercept)  -21.68727 2425.75982  -0.009 0.992867    
years_active   0.09013    0.03646   2.472 0.013441 *  
imdb_credits  -0.01791    0.01018  -1.760 0.078337 .  
family         2.12217    0.55867   3.799 0.000146 ***
education1    19.25220 2425.75968   0.008 0.993668    
education2    19.67544 2425.75964   0.008 0.993528    
education3    35.60365 3295.98625   0.011 0.991381    
education4    39.67097 3198.57518   0.012 0.990104    
wins          -0.97875    0.38555  -2.539 0.011130 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 160.176  on 117  degrees of freedom
Residual deviance:  76.005  on 109  degrees of freedom
AIC: 94.005

Number of Fisher Scoring iterations: 17
```

Table 4

Fitted logistic regression model confusion matrix

```
Confusion Matrix and Statistics

          Reference
Prediction  0  1
         0 62 10
         1  7 39

               Accuracy : 0.8559
                 95% CI : (0.7794, 0.9138)
    No Information Rate : 0.5847
    P-Value [Acc > NIR] : 1.713e-10

                  Kappa : 0.7007

 Mcnemar's Test P-Value : 0.6276

            Sensitivity : 0.8986
            Specificity : 0.7959
         Pos Pred Value : 0.8611
         Neg Pred Value : 0.8478
             Prevalence : 0.5847
         Detection Rate : 0.5254
   Detection Prevalence : 0.6102
      Balanced Accuracy : 0.8472

       'Positive' Class : 0
```

Test Predictions

A. Prediction for Vilma Santos

Vilma Santos

```
{r}
vilma <- data.frame(
  years_active = 62,
  imdb_credits = 231,
  family = 2,
  education = as.factor(4),
  wins = 18
)
vilma_prob <- predict(model_no_outliers1, vilma, type = "response")
vilma_class <- ifelse(vilma_prob > 0.45, 1, 0)
vilma_class
```

```
1
1
```

B. Prediction for a Hypothetical HS Graduate Bong Revilla

Bong Revilla (assume only HS graduate)

```
{r}
bong_hs <- data.frame(
  years_active = 42,
  imdb_credits = 100,
  family = 6,
  education = as.factor(1),
  wins = 2
)
bong_hs_prob <- predict(model_no_outliers1, bong_hs, type = "response")
bong_hs_class <- ifelse(bong_hs_prob > 0.45, 1, 0)
bong_hs_class
```

```
1
1
```

C. Prediction for Coco Martin

Coco Martin

```
{r}
coco <- data.frame(
  years_active = 24,
  imdb_credits = 66,
  family = 0,
  education = as.factor(2),
  wins = 1
)
coco_prob <- predict(model_no_outliers1, coco, type = "response")
coco_class <- ifelse(coco_prob > 0.45, 1, 0)
coco_class
```

```
1
0
```

D. Prediction for Willie Revillame

Willie Revillame

```
{r}
wowowie <- data.frame(
  years_active = 38,
  imdb_credits = 38,
  family = 0,
  education = as.factor(0),
  wins = 0
)
wowowie_prob <- predict(model_no_outliers1, wowowie, type = "response")
wowowie_class <- ifelse(wowowie_prob > 0.45, 1, 0)
wowowie_class
```

```
1
0
```

E. Prediction for Jolo Revilla

Jolo Revilla

```
{r}
jolo <- data.frame(
  years_active = 19,
  imdb_credits = 19,
  family = 2,
  education = as.factor(2),
  wins = 0
)
jolo_prob <- predict(model_no_outliers1, jolo, type = "response")
jolo_class <- ifelse(jolo_prob > 0.45, 1, 0)
jolo_class
```

```
1
1
```

F. Prediction for Ejay Falcon

Ejay Falcon

```
{r}
ejay <- data.frame(
  years_active = 17,
  imdb_credits = 44,
  family = 0,
  education = as.factor(2),
  wins = 0
)
ejay_prob <- predict(model_no_outliers1, ejay, type = "response")
ejay_class <- ifelse(ejay_prob > 0.45, 1, 0)
ejay_class
```

```
1
0
```

G. Prediction for Daniel Fernando

Daniel Fernando

```
{r}
danf <- data.frame(
  years_active = 32,
  imdb_credits = 110,
  family = 0,
  education = as.factor(2),
  wins = 1
)
danf_prob <- predict(model_no_outliers1, danf, type = "response")
danf_class <- ifelse(danf_prob > 0.45, 1, 0)
danf_class
```

```
1
0
```