The Secret Ingredients of Love:

Predicting Success of Celebrity Romance

Wu Linhan[1], Xu Yawan[2] and Edelmann Gioia[3]

1. Wee Kim Wee School of Communication and Information, Nanyang Technological University

2. School of Arts and Science, McGill University

3. School of Management and Law, Zurich University of Applied Sciences

Acknowledgements

This research was carried under the guidance and supervision of professor Poong Oh.

We would like to thank him for providing insight and expertise that greatly assisted the research.

Abstract

This study proposes the use of a Naïve Bayes classification model in predicting the likelihood of success of a given celebrity romantic relationship, with success pointing to the partners getting engaged or married. Bringing computational methods into relationship predictions, this study adds a new perspective to previous works that made primary use of qualitative methods. With individual couples as units of analysis, the model is based on features that are shared by both partners in a given relationship. The underlying assumption is that these identified features contribute to a relationship's success and can therefore be applied to predict its future outcome. Having tested different combinations of features with our dataset of 90 celebrity couples, the best result was attained by the model with binary observations of the couples' relationship durations as well as differences in age, height, ethnicity and number of previous relationships. With the model's predictive power being moderate, further research is necessary for improvements. On the technical level, potential improvements could be achieved by the use of a larger dataset or the use of other classifiers such as Support Vector Machines. On the concept level, improvements could be achieved through the integration of additional features such as the presence of children or differences in political views. On top of playing a role in proliferating popular culture and swaying celebrity romances, the findings and model from this study could serve as a basis that empowers everyday couples to make better-informed decisions based on their shared characteristics.

Keywords: *celebrity, romance, relationship, prediction, Naïve Bayes*

The Secret Ingredients of Love: Predicting Success of Celebrity Romance

The forming and dissolution of celebrity romantic relationships have always served as newsworthy hearsay in our private circles. Emotions run high and the floodgates of gossips overflows whenever our favorite musicians, actors or actresses break up, or found a new mysterious soulmate. Besides having news value, changes in celebrity relationships were discovered by *Matches.com*'s *Singles in America* study to be more similar to those of everyday couples than previously believed (Match.com, 2012). With 55% of men get over a breakup in under three months, behaviors like this mirrors those of celebrities. This suggests celebrity couples could be used as a proxy to draw inferences and study romantic relationships between everyday people like you and me. That is why this study set out to investigate the question "How likely is it for a particular celebrity couple to stay together?"

From adolescents to married couples, previous studies in the area of romantic relationships have been largely focused on understanding the factors behind successful and failed romances of everyday people. A study by Ireland et al. (2010) discovered that couples that share more similarities in terms of language use are more likely to stay together, indicating that language reflects implicit interpersonal behaviors. Another study by Vohs et al. (2010) studied the effects of self-control as a factor of romantic success, discovering that couples with high relationship satisfaction were also of high self-control. While informing us about the notable factors of successful relationships, most previous studies had neither undertaken a computational approach to predict relationship successes nor investigated celebrity relationships.

With a primary objective of computationally predicting the success of romantic relationships with celebrity couples as units of analysis, our study fills in the gap within this field of research. The secondary objective of this study is to identify the salient factors of a lasting romance. Success, in the case of this study, refers to a couple who got engaged or married.

Being able to predict the future of celebrity relationships would not only bring about significant news value for gossip columns and tabloids, but also quench the curiosity of die-hard fans, who would go about spreading the news to friends and family. Such predictions would in turn play a role in proliferating popular culture and might even sway the decisions of celebrities to date one another.

Method

1. Formulation of research challenge

Our research challenge is to predict the likelihood of success of a given celebrity romantic relationship, with success pointing to the partners getting engaged or married. As this challenge deals with the assignment of a couple to either the class of *failed* (break-up) or *success* (stay-together) based on numerous shared features, it can be best solved using a classification model (Rebala, Ravi & Churiwala, 2019). As a supervised machine learning method, classification models require sets of training data with labelled classes from which the model learns (Rebala et al., 2019). In our case, the training data comprises both failed and successful celebrity relationships with which we hope to develop a model that predicts, with high accuracy, if a given celebrity couple is going to stay together. Furthermore, we aim to learn about the features in our model that are more important in accurately predicting a celebrity couple's common destiny.

With individual pairs of couples being the unit of our analysis, we identified the following six features based on the availability of information as well as the underlying assumption that they do have a tangible impact on a couple’s relationship. The features will be further explained in the data collection section.

Table 1: Features used to predict class category

Features	Class Category
Relationship Duration	
Difference in Number of Previous Relationships	
Age Difference	Success (Stay Together)
Height Difference	
Religion	Failure (Break Up)
Ethnicity	
Butterfly Score	

For our classification model, we chose the Naïve Bayes classifier as it offers a reliable way to gain a basic first insight into the research challenge while still being reasonably precise (Cichosz, 2015). It allows us to quickly try out several approaches and learn about the prediction quality of our features before working with more complex and expensive classifiers (Cichosz, 2015).

Using Naïve Bayes, our research problem can be formulated as follows:

$$P(stayTogether \mid values\ in\ features = \frac{P(values\ in\ features \mid stayTogether)\ P(stayTogether)}{P(values\ in\ features)}$$

$$P(breakUp \mid values\ in\ features) = \frac{P(values\ in\ features \mid breakUp)\ P(breakUp)}{P(values\ in\ features)}$$

As can be seen in the above equations, Naïve Bayes makes use of the fact that the probability of belonging to a class *C* given values in the features *x* can be calculated based on the prior or marginal probability and the inverse conditional probability (Cichosz, 2015). For our study, this implies that we can calculate both the probability of a celebrity couple staying together as well as breaking up based on their shared features. The larger of the two probabilities determines a couple's class category and hence, the model's prediction for that couple.

2. Data Collection

The data for this study was sourced from the website *whosdatedwho.com*, which has information on relationship histories and biographies of over 50,000 celebrities. The information of 90 pairs of randomly selected celebrity couples, where both partners are celebrities, were manually harvested. Half of them have "successful" relationships, where the partners are currently married or engaged, and the other half "failed", where the partners were in a relationship for over three months before splitting up (break up or divorce).

Table 2: Descriptions of collected and transformed features

Features Collected	Description	Transformed Features	Description
Duration of Relationship	Entire relationships not just marriage or engagement duration	DurRe	Duration of Relationship
Number of Previous Relationships	Including previous marriages and encounters like flings	rel_diff	Difference in number of previous relationships *(male - female)*
Age	Ages of both partners	age_diff	Difference in partners' ages *(male - female)*
Height	Heights of both partners	height_diff	Difference in partners' heights *(male - female)*
Religion	Religions of both partners	same_religion (boolean)	If both religions are the same
Ethnicity	Ethinicities of both partners	same_ethnicity (boolean)	If both ethnicities are the same
		butterfly_score	The propensity of a couple to break up. Given by the sum of both partners' number of relationships divided by their ages (to account for older partners having more relationships)

Six pieces of information were collected from each partner's biography page and the differences in continuous variables *number of previous relationships*, *age* and *height* were calculated by deducting the values for the female from that of the male. In the case of same sex couples, we took the absolute values.

Butterfly_score is a value that was developed to gauge the propensity of a couple to break up based on the regularity of the partners' relationships. However, the underlying assumption is that previous marriages and encounters such as flings are all previous relationships, which might result in unrepresentative scores for celebrities who had more encounters than actual relationships.

To facilitate further formulation of our predictive model, the continuous features, *rel*_diff, age*_diff*, *height_diff* and *butterfly_score,* were then transformed into categorical features of 2 and 3 categories by equal proportions.

3. Data Analysis

Test for feature independence. R is the computational software used for all our data analysis. As Naïve Bayes classifiers assumes independence among input features, the *psych* package was used to plot and glance through the correlation across pairs of input features.

Data splitting. The dataset was then split into training and test sets with a 70-30 train-test split using the *caret* package for its stratified sampling. It maintains equal proportions of the dependent classes (*success* and *failed* relationships) in both training and test sets, which ensures congruence and internal validity.

Model Training, improvement and results. In the following stage, the Naïve Bayes classification model is trained. In place of manually writing the Naïve Bayes algorithm, the *naivebayes* package was chosen to facilitate much faster iterations of our classification models when identifying the optimal set of features to use. The statuses of relationships are the target classes. Results of the model at each iteration were quickly detailed using the *caret* package before deciding if any changes were to be made to the combination of features. Due to our dataset being limited in quantity, repeated 10-fold cross validation is also employed in attempts to further improve model performance.

Results

1. Statistical Analysis

Input features are not strongly correlated. The test for feature independence using the *psych* package returned low correlation coefficients across our input features, with the highest being -0.44 (on a scale of -1 to 1) between *butterfly_score* and *durRe* (duration of relationship). This indicates that our results from a Naïve Bayes classification model would be more applicable to real life than in the case of having dependent features because the model itself assumes independence between input features. This validates our choice of selecting Naïve Bayes.

Significant difference in butterfly_score between Success and Failed couples. With a t-statistic of 3.56 and a negligibly small p-value, couples with failed relationship have significantly higher *butterfly_scores* as compared to couples with successful relationships. With age accounted for, this indicates that failed couples had more relationships, and might suggest that they are more likely to break up and change partners.

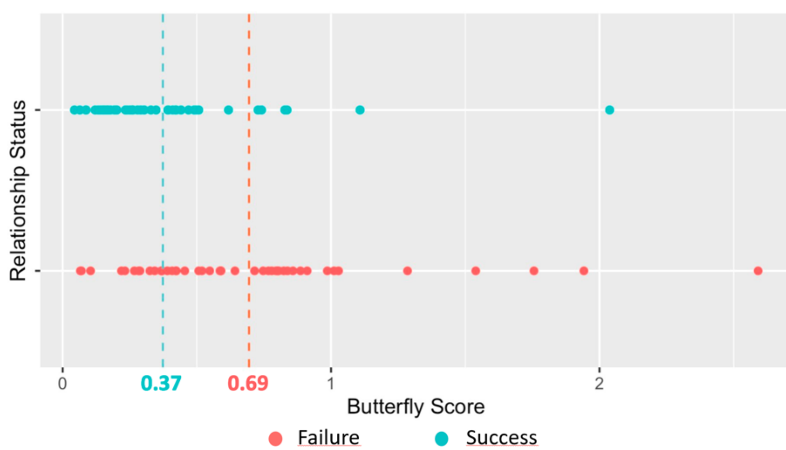

Figure 1: Butterfly Score by Relationship Status

2. Naïve Bayes Classification Models

Naïve Bayes classification models were trained with different combinations of features to develop the most accurate model. Table 3 details the accuracies of the models.

Table 3: Performances of Classification Models 1 to 3

Model	Features Used	Accuracy
1	all original features (with continuous inputs)	57.7%
2	all features (with continuous inputs converted into 3 categories)	61.5%
3	all features (with continuous inputs converted into 2 categories)	69.2%

Our team believed that continuous inputs are more precise and would therefore contribute to better model performances than when we decrease available information by categorization. However, the opposite results were unexpectedly true. In addition, when we further reduced the available information by having two categories for each continuous feature, the accuracy increased even more. This might be due to our dataset being too small to train a model with more categories.

Further improvements were made with Model 3 as the basis by removing input features. Table 4 details the accuracies of the following models.

Table 4: Performances of Classification Models 4 to 7

Model	Features Used	Accuracy
4	Model 3 without Duration of Relationship	53.8%
5	Model 3 without Religion	73.1%
6	Model 3 without Religion & *butterfly_score*	76.9%
7	Diff in Age, Height & Duration of Relationship	76.9%

duration of relationship is a useful predictor as its removal resulted in a drastic decrease in accuracy. Retracing our data collection steps, our team realized that *religion* had over 50% of *NA* instances that could have introduced noise into the dataset. That was why the removal of the *religion* feature increased accuracy, setting a new basis for further improvement. In Model 6, we removed *religion* and *butterfly_score*, increasing the accuracy once again. This was contrary to our prior belief that *butterfly_score* is a good indicator of the propensity of a couple to break up, as supported by the significant difference between the scores of successful and failed relationships.

Further removal of features did not result in a change of model performance until only three features remained, from which point the performance decreases. Models with 10-fold cross-validation unexpectedly resulted in the same performances. Model 6 was selected as the final model for its better performance while retaining more information than later models.

Table 5: Performance of Classification Model 6

Accuracy			
76.9%			
Predicting Success		Predicting Failure	
Precision	0.8182	Precision	0.8462
Recall	0.6923	Recall	0.7333
F-Measure	0.7500	F-Measure	0.7857

Table 6: Correct predictions made with probability of more than 80%

same_ethnicity	height_diff_cat	age_diff_cat	rel_diff_cat	durRe_cat	reStatus	pred	prob
TRUE	TRUE	TRUE	FALSE	FALSE	Y	Y	0.8491220
TRUE	FALSE	TRUE	FALSE	FALSE	Y	Y	0.9069675
TRUE	TRUE	TRUE	TRUE	FALSE	Y	Y	0.8148291
TRUE	FALSE	FALSE	FALSE	FALSE	Y	Y	0.8532712
TRUE	FALSE	FALSE	FALSE	FALSE	Y	Y	0.8532712
TRUE	FALSE	FALSE	FALSE	FALSE	Y	Y	0.8532712
TRUE	FALSE	FALSE	TRUE	FALSE	Y	Y	0.8197212

While our model is better overall at predicting for relationship failures, correct predictions made with over 80% probability were all for successes, as shown in Table 6. We noticed that the most salient conditions that lead to a "successful" relationship is that of having the same ethnicity. Following that, when the duration of relationship is above 3 years (expressed as *FALSE*), the probability of "success" is higher. To a smaller extent, "success" probability is higher where the male partner is more than 14cm taller than the female partner (expressed by *FALSE*). Lastly, the same could be said for relationships when the male partner has over 3 previous relationships more than the female partner (expressed by *FALSE*).

Discussion

1. Theoretical and Practical Consequences

Insights from our research could potentially help a wide range of people like researchers, the paparazzi, and even everyday individuals. For instance, sociologists might be able to draw inferences from our results and make progress in their investigations in human social behaviors. Gossip websites might use our model to predict the relationship outcome of a celebrity's rumored relationship. Finally, everyday individuals can benefit from our study by predicting their relationship outcome and finding out if they are with the right partner.

2. Limitations and Potential Bias

As our results have shown, our model only performed moderately. There are in fact several limitations in our research that hindered its performance. We will examine each of these below.

Census Problem. Our study deals with data that might be outdated by decisions of any "successful" couple to get divorced, or any previously "failed" couple to get together again. We currently have no method to peg the validity of our research against such future changes.

Sample Size. One important limiting factor might be our small sample size. Although the sample might be representative of the population of celebrities, as it was randomly chosen, the data of just 90 couples might not be enough to train our model. However, we do believe that there is a threshold at which an increase in information will no longer result in higher performance, as we hypothesize that the nature of human relationship is too complex to be modelled perfectly.

Sample quality. We should have imposed more criteria when collecting the data. Our dataset included homosexual couples, as we did not think that the gender of the partner mattered. However, when we calculated the age, height, and number of relationship differences, we arbitrarily chose to subtract the females' values from those of the males' to account for not just the absolute difference between the two partners, but also to see whether there a gender bias existed. Of course, we could not impose an order for the homosexual couples, and therefore, compromising the quality of the dataset.

Lack of available and/or reliable data. Although we had access to the large celebrity database via *whosdatedwho.com*, the information on the website was not always accurate, in particular, for the number of relationships reported. For instance, a popular celebrity might have a lesser known spouse, whom we have less reliable and available information on. Often, it would be shown that their number of relationships is one, while that of their popular counterpart is in the double-digits. However, this might be simply of that we don’t know enough about their private life. Since we cannot make the distinction between a lack of data and real data, we might have introduced biases in the *butterfly_score*.

Subjective definition of success. What determines a successful relationship is subjective and ambiguous. We have defined success as couples who are currently engaged or married. However, we have not set further criteria for the length of marriage or any other qualitative characteristics. Furthermore, without primary research, it is very difficult to determine the quality of a relationship. As such, although there is no guarantee that a married couple is going through a more “successful” relationship than a nonmarried couple, we deemed marital status to be the one and only criteria for success in our research.

3. Potential Improvements and Future Research

We believe that our research could be improved in many ways. The first would be to change our method of data collection. Rather than gathering celebrities’ data from an online source, a more reliable method would be conducting primary research. One benefit of doing so is that first-person report of information eliminates the chances that a lesser known celebrity has less accurate data. Furthermore, we could extract additional features by asking participants specific questions that we deem relevant, such as how many children they have with their partner, how they would rank their relationship quality, how long ago the respondent starts to live with partner. All of which were hard to access through secondary research.

One potential improvement of the model could be making it applicable for same-sex couples as well. For that, we would need a gender-neutral model, where the quantitative differences (eg., height difference) would not take into account gender biases.

Similarly, another way to build on the model could be to have an ethnicity-specific model to observe whether there are differences between relationships in Asians versus in Caucasians.

On the computational side, it would be interesting to experiment with different classification algorithms, such as logistic regression and SVM. These two models have given Lou and Yang the highest accuracies in their research (Lou and Yang, 2016).

Conclusion

Our research aimed to shed light on the complexity of romantic relationships by studying celebrities' successful and failed relationships. We successfully built a Naïve Bayes model that predicts celebrity break up probability with a F-measure score of 0.786. Although its performance is only moderate, we theorize that it is partly due to the complex nature of human relationships. Hence, we believe that our model could still be a valid basis on which future research take place.

While our study did offer insights into the mystery of relationships, we are puzzled by some of the results that we obtained. For instance, why is it that a relationship is more likely to succeed when the male has had more past experiences than the female? Is this due to a bias in the dataset? Future research is necessary to answer these questions.

References

Cichosz, P. (2015). *Data Mining Algorithms: Explained Using R.* Hoboken, NJ:John Wiley&Sons, Ltd.

In Hollywood it seems everyone's dated everyone! (n.d.). Retrieved from https://www.whosdatedwho.com/.

Lou, J., & Yang, H. (2016). Have You Met The 1? A Machine's Approach to Human Relationships. Retrieved on 21 October 2019 from http://cs229.stanford.edu/proj2016spr/report/007.pdf

Khan, R. (n.d.). RPubs – Naive Bayes Classifier: Theory and R example. Retrieved on 21 October 2019 from https://rpubs.com/riazakhan94/naive_bayes_classifier_e1071

LE, B. , DOVE, N. L., AGNEW, C. R., KORN, M. S. and MUTSO, A. A. (2010), Predicting nonmarital romantic relationship dissolution: A meta‐analytic synthesis. Personal Relationships, 17: 377-390. doi:10.1111/j.1475-6811.2010.01285.x

Prabhakaran, S. (2018). How Naive Bayes Algorithm Works? [Blog Post]. Retrieved on 21 October 2019 from https://www.machinelearningplus.com/predictive-modeling/how-naive-bayes-algorithm-works-with-example-and-full-code/

Rebala, G., Ravi, A., & Churiwala, S. (2019). Classification. In G. Rebala, A. Ravi, & S. Churiwala (Hrsg.), *An Introduction to Machine Learning* (S. 57–66). https://doi.org/10.1007/978-3-030-15729-6_5

Ireland, M. E., Slatcher, R. B., & Eastwick, P. W. (2011). Language Style Matching Predicts Relationship Initiation and Stability. *Association for Psychological Science*, *22*(I), 39–44. Retrieved on 21 October 2019 from https://journals.sagepub.com/doi/abs/10.1177/0956797610392928

Vohs, K. D., Finkenauer, C., & Baumeister, R. F. (2011). The Sum of Friends' and Lovers'Self-Control Scores Predicts RelationshipQuality. *Social Psychological AndPersonality*

Science, *2*(2), 138–145. Retrieved on 21 October 2019 from https://journals.sagepub.com.remotexs.ntu.edu.sg/doi/pdf/10.1177/1948550610385710