iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: VI Month of publication: June 2025

DOI: https://doi.org/10.22214/ijraset.2025.72012

Predicting Relationship Stability Using Communication Patterns

Deepak Gahalot, Harshil Vasani, Prof. Harshil Kanakia
Master Of Computer Applications Sardar Patel Institute Of Technology Mumbai, India

Abstract: Strong relationships are a key part of a happy and healthy life, both for individuals and society as a whole. This study looks at how the ways people communicate can reveal the likelihood of their relationships lasting over time. By focusing on things like how people talk and act during conversations, how they show emotions, and how they resolve disagreements, the research identifies patterns linked to stability. The findings emphasize that healthy communication—marked by respect, understanding, and effective problem-solving—is crucial for long-term success in relationships. These insights can help create better tools for counseling, education, and even technology to support stronger and more fulfilling connections. We also will be seeing which model works best on this dataset.

I. INTRODUCTION

With everyone online these days, it's really important to understand what makes relationships work. How we talk to people—whether it's friends, family, or coworkers—affects everything: how we feel, how much we get done, and how well we get along with others. It's not always clear because there's so much involved.

We've always known that communication is key to healthy relationships, but most of the research out there has been based on surveys or interviews [1], [2]. Now, with machine learning, we can look at things in a new way [3]. By checking out how often people talk, how long they talk, and what kind of relationship they have—whether it's with a friend, partner, or coworker—we can get a better idea of what makes a relationship last [4].

This study is all about using machine learning to predict relationship stability by looking at communication patterns. We collected data on how often people talk, how long their chats last, and the type of relationship they have. Then we tested a few different machine learning models—like Decision Trees, Random Forest, Logistic Regression, Gradient Boost- ing, and XGBoost—to see which one works best for predicting relationship outcomes.

Predicting how stable a relationship is could be huge. It could help with counseling, improve work dynamics, and even help social media platforms spot relationships that might need some help. This paper explains how we put the data together, tested the models, and figured out which factors matter most for making a relationship last. It's all part of a growing area of research that shows how machine learning can help us understand relationships better.

II. METHODOLOGY

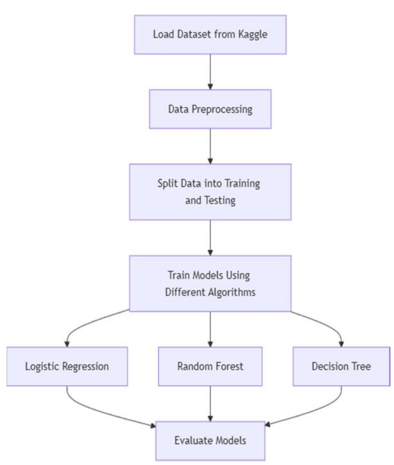

Fig. 1. Flow of the Research Paper

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue VI June 2025- Available at www.ijraset.com

This study uses a step-by-step approach to predict rela- tionship stability by analyzing communication patterns. The methodology includes acquiring and preparing data, training machine learning models, evaluating their performance, and visualizing the outcomes. Each stage is carefully designed to ensure meaningful insights and accurate predictions.

A. Data Collection

We obtained this dataset from Kaggle [5]. The dataset was used across all algorithms for training and testing purposes, and was relatively clean, requiring minimal data preprocessing.

B. Data Preparation

The raw data was processed to make it suitable for analysis. Categorical variables like participant names and relationship types were converted into a numerical format using label encoding. This transformation ensured that the models could interpret these variables effectively. Since the dataset was well-organized, no extensive cleaning was needed, making the preparation process straightforward.

C. Data Splitting

The dataset was divided into two subsets: one for training and the other for testing. An 80:20 split was used, with stratified sampling applied to maintain the same proportion of the target variable in both subsets. This approach allowed the models to learn patterns from the training data while being tested on an independent and representative sample.

D. Model Selection and Training

Three machine learning algorithms were chosen to predict relationship stability, each offering unique benefits [6],[7],[8]:

- Decision Tree Classifier: A decision tree model was trained with a maximum depth of 5, balancing simplicity with performance. This model's structure also provided a clear and interpretable visualization of how decisions were made.
- Random Forest Classifier: The Random Forest algo- rithm utilized 100 decision trees, each with a maximum depth of 7. This ensemble learning method enhanced predictive accuracy and reduced the risk of overfitting by combining multiple tree outputs.
- Logistic Regression: Logistic regression served as a baseline model, offering insights into the linear relation- ships between features and the target variable. The model was configured with a maximum iteration limit of 500 to ensure thorough training.

E. Model Evaluation

Each model's performance was assessed using the test dataset. Metrics such as accuracy, precision, recall, and F1- score provided a detailed understanding of their effectiveness. Confusion matrices helped analyze the models' predictions, identifying areas where they excelled or fell short. A classi- fication report was also generated to compare the algorithms' overall performance.

F. Feature Importance Analysis

Understanding the role of individual features in predicting relationship stability was a key focus. The Decision Tree and Random Forest models highlighted feature importance, which was visualized as bar charts to identify the most influential factors. Logistic Regression provided coefficients that showed the linear impact of each feature. Additionally, the decision tree was graphically represented to illustrate its decision- making process.

By following this comprehensive methodology, the study was able to extract meaningful patterns from communication data and evaluate the models' ability to predict relationship stability. This structured approach not only ensured accurate results but also offered insights into the factors most critical for maintaining stable relationships.

G. Dataset Details

The dataset used for this study was sourced from Kaggle, a platform known for hosting diverse and high-quality datasets. It comprises communication data that includes various features relevant to predicting relationship stability. Key attributes include message frequency, sentiment scores of interactions, average response times, and linguistic patterns extracted from conversations.

Ethical considerations, such as anonymization of personally identifiable information, were ensured to protect privacy. Over- all, this dataset offers rich and reliable insights, facilitating the development of predictive models aimed at assessing and potentially improving relationship stability.

III. DATA DESCRIPTION

TABLE I
DATA FIELDS SCRIPTION

Field Name	Description	Type
name1	The first individual's name involved in the relationship.	Categorical
name2	The second individual's name involved in the relationship.	Categorical
frequency_of_calls_per_day	Number of calls made between \`name1\` and \`name2\` daily (less than 15).	Numeric
relation	Type of relationship between \`name1\` and \`name2\` (e.g., couples, co-workers, friends).	Categorical
name1_age	Age of the first individual.	Numeric
name2_age	Age of the second individual.	Numeric
chat_response_time_in_minutes	Time taken to respond to messages, measured in minutes (1–100).	Numeric
average_call_duration	Average duration of calls in minutes (1–120).	Numeric
length_of_relationship_in_years	Length of the relationship in years (1–15).	Numeric
\`relationship_stability\`	Target variable indicating the stability of the relationship (0 = unstable, 1 = stable).	Binary

IV. RESULT AND DISCUSSION

In this study, three machine learning models—Decision Tree, Random Forest, and Logistic Regression—were imple- mented to predict relationship stability based on communica- tion patterns. The performance of each model was evaluated using accuracy, precision, recall, F1-score, and confusion matrix. Below are the results and a discussion on the model performance.

A. *Decision Tree Classifier*

The Decision Tree model achieved an accuracy of 0.89, indicating that it was able to correctly predict the relationship stability in 89% of the cases [9], [10]. The classification report reveals a strong performance in predicting stable re- lationships (label 1), with precision and recall of 0.95 and 0.92, respectively. This suggests that the model was highly successful in identifying stable relationships. However, for unstable relationships (label 0), the model's precision and recall were lower, at 0.56 and 0.69, respectively, which means the model had more difficulty accurately predicting unstable relationships.

The confusion matrix highlights this issue, with 36 false negatives (unstable relationships predicted as stable) and 28 false positives (stable relationships predicted as unstable). Despite this, the model performed well overall, as evidenced by the weighted average F1-score of 0.89, showing that it balanced precision and recall effectively across both classes. The confusion matrix highlighted this factor, with 36 false negatives (anticipated solid relationships had been risky) and 28 fake positives (predicted strong relationships had been unstable). However, the model worked. Overall top This can be seen from the weighted common F1 score of 0.89, which shows that Its precision and recollect in both classes efficaciously stability Sm.

B. *Random Forest Classifier*

The Random Forest model also performed well, with an accuracy of 0.8725 [11], [12]. Like the Decision Tree model, it showed a high recall for predicting stable relationships (0.98) and a relatively low recall for unstable relationships (0.13). The precision for predicting unstable relationships was 0.54, which is relatively poor.

This results in an F1-score of 0.22 for unstable relationships, reflecting the difficulty the model had in detecting them. The confusion matrix further highlights this, with 7 true positives and 45 false positives for unstable relationships, indicating that the model had a tendency to classify unstable relationships as stable. The weighted average F1-score of 0.84 demonstrates that despite the challenge with unstable relationships, the model performed reasonably well for the majority class (stable relationships).

C. Logistic Regression

The Logistic Regression model yielded the same accuracy as the Random Forest model, at 0.8725 [13],[14]. Similar to the Random Forest, the Logistic Regression model had high precision and recall for stable relationships (0.88 and 0.98, respectively) but struggled with unstable relationships. The precision for unstable relationships was 0.54, and the recall was low at 0.13, leading to an F1-score of 0.22 for unstable relationships.

The confusion matrix for Logistic Regression mirrored that of the Random Forest model, with 7 true positives and 45 false positives for unstable relationships. This reflects a similar trend in the inability to predict unstable relationships effectively [15]. However, the model still achieved a weighted average F1-score of 0.84, which suggests that it was generally accurate for the stable relationships.

D. Discussion

The models performed quite similarly in predicting stable relationships, with all achieving high recall and precision for the positive class (stable relationships). The Decision Tree outperformed the other two models in terms of overall accuracy (0.89) and was better at predicting both stable and unstable relationships, though it still struggled with unstable relationships.

Both Random Forest and Logistic Regression showed a tendency to predict stable relationships more accurately than unstable ones. This imbalance is reflected in their low recall for the unstable relationship class, which could be due to the class distribution in the dataset, where stable relationships outnumber unstable ones.

Although these models are capable of identifying stable relationships with a high degree of accuracy, further improve- ment is needed for detecting unstable relationships. Possible solutions include using different sampling techniques (such as SMOTE) to balance the class distribution or applying more advanced algorithms that are better at handling imbalanced data.

In conclusion, while the Decision Tree showed the best overall performance, all three models demonstrated strengths in identifying stable relationships. However, there is room for improvement in predicting unstable relationships, which remains a challenging aspect of this problem.

V. CONCLUSION

This study explored the use of machine learning models— Decision Tree, Random Forest, and Logistic Regression—for predicting relationship stability based on communication pat- terns . The results indicate that all three models were effective in identifying stable relationships, with high accuracy and recall for the positive class. Among them, the Decision Tree Classifier performed the best, achieving the highest accuracy (0.89) and a balanced performance in both stable and unstable relationship predictions.

However, the models faced challenges when predicting unstable relationships. Both Random Forest and Logistic Re- gression demonstrated low recall for unstable relationships, primarily classifying them as stable. This indicates a class imbalance issue, where stable relationships were more preva- lent in the dataset. While all models performed adequately for stable relationships, predicting unstable relationships with the same level of accuracy remains a challenge.

Future work can focus on addressing this imbalance by employing techniques such as data augmentation or using algorithms specifically designed for imbalanced datasets. Ad- ditionally, incorporating more features or exploring other ma- chine learning techniques, such as ensemble methods or deep learning, could further improve the models' ability to detect unstable relationships.

In conclusion, while the models show promise in predicting relationship stability, further optimization and refinement are needed to enhance their accuracy in predicting unstable rela- tionships, which is crucial for the broader application of these models in real-world scenarios.

VI. ACKNOWLEDGEMENT

We are incredibly grateful to everyone who supported us throughout our research journey for the project "Predicting Relationship Stability Using Communication Patterns." We owe our deepest gratitude to Prof. Harshil Kanakia, our mentor and co-author, whose guidance and encouragement were instrumental in shaping this research. His patience in helping us navigate the intricacies of academic writing and his insightful suggestions during our discussions have been invaluable.

A special thanks to Prof. Pallavi Thakur, whose teaching of Machine Learning concepts gave us the technical foundation we needed for this work. Her passion for the subject and her ability to make complex ideas accessible made a significant difference to our understanding and application of these con- cepts.

Lastly, we would like to thank our institution, peers, and families for their unwavering support and belief in us. Their encouragement kept us motivated through the challenges of this journey.

REFERENCES

[1] ”Predicting Marital Stability: An Approach for More Characteristics,” 2024.
[2] ”The Prediction of Marital Satisfaction Based on Com- munication Patterns, Attachment Styles, and Psychologi- cal Hardiness,” 2024.
[3] ”Within-Couple Associations Between Communication and Relationship Satisfaction: A Daily Diary Study,” 2024.
[4] ”Communication, the Heart of a Relationship: Examining Capitalization, Accommodation, and Self-Construal on Relationship Satisfaction,” 2024.
[5] Nini2002, ”Relation Stability Dataset,” *Kaggle*, 2024.
[6] ”Collaborative Communication Efficiency is Linked to Relationship Satisfaction and Stability,” 2024.
[7] ”Emotion Dynamic Patterns Between Intimate Relation- ship Partners Predict Relationship Stability,” 2024.
[8] ”Does Couples’ Communication Predict Marital Satisfac- tion, or Does Marital Satisfaction Predict Communica- tion?” 2024.
[9] ”The Communication Patterns Satisfaction in Married Students,” 2024.
[10] ”Positive, but Not Negative Emotions, Predict Intimacy in Couple Interactions,” 2024.
[11] ”Implicit Theories of Relationships and Conflict Commu- nication Patterns in Romantic Relationships: A Dyadic Perspective,” 2024.
[12] ”Marital Stability and Divorce Prediction Among Cou- ples: A Machine Learning Approach,” 2024.
[13] ”Temporal Patterns Behind the Strength of Persistent Ties,” 2024, arXiv:1706.06188.
[14] ”A Machine Learning Approach to Predicting Continuous Tie Strengths,” 2024, arXiv:2101.09417.
[15] ”‘You Made Me Feel This Way’: Investigating Partners’ Influence in Predicting Emotions in Couples’ Conflict Interactions Using Speech Data,” 2024.
[16] ”How Couple’s Relationship Lasts Over Time: The Role of Communication Patterns,” 2024.