KULLIYYAH OF INFORMATION AND COMMUNICATION TECHNOLOGY

CSCI 4343 DATA SCIENCE

SEMESTER 2, 2024/2025

GROUP PROJECT REPORT:

IDENTIFYING FACTORS AFFECTING MUSLIM MARRIAGE SUCCESS USING DATA ANALYSIS

GROUP: 6

NAME	MATRIC NO	Email
MUHAMMAD NAZRIN BIN JAMIL	2219905	mnazrinj100@gmail.com
MUHAMMAD LUQMAN BISTHAMY BIN SHAM ZULANEY	2216289	luqmanbisthamy991@gmail.com
WAN AHMAD HAFIZUDDIN BIN WAN HUSSIN	2217135	wanhafizuddin0657@gmail.com
UMAIR BIN MOHD YAZIZ	2123467	umairmohdyaziz@gmail.com

INSTRUCTED BY:

DR. RAINI BINTI HASSAN

TABLE OF CONTENTS

ABSTRACT

Islam places a high value on marriage because it provides companionship, emotional support, and a basis for societal stability. However, divorce rates among Muslim spouses have alarmingly increased in Malaysia in recent years. This expanding tendency emphasises how crucial it is to comprehend what makes a marriage succeed.

Utilizing data analytics and machine learning, this study seeks to identify the key factors that contribute to the success of Muslim marriages in Malaysia. The study will investigate how factors such as communication, religious commitment, financial stability, and external influences impact marital happiness by combining findings from a primary survey with secondary data from sources like Kaggle.

The project will use predictive analytics, such as sentiment analysis, categorisation models, and statistical correlation, to find significant patterns and trends in order to do. A sentiment analysis dashboard, a prediction model, and concise, useful information for couples, marital counselors, religious leaders, and legislators are the objectives. The ultimate goal of these studies is to aid in the creation of more efficient, evidence-based plans to improve marriages and lower divorce rates among Muslims in Malaysia.

INTRODUCTION

In recent years, the rising divorce rates among Muslim couples in Malaysia have become a growing social concern. In Islam, marriage is meant to foster emotional, spiritual, and social well-being, yet many couples today face challenges such as financial pressures, poor communication, and changing societal expectations that can lead to marital instability and separation.

This study seeks to explore the key factors that influence the success of Muslim marriages in Malaysia. Positioned within the field of business-related social analytics, it uses data science methods to uncover meaningful patterns and offer practical recommendations for couples, marriage counselors, and policymakers. The research aims to develop a predictive analytics model that can identify traits and conditions linked to lasting, successful marriages based on data gathered from Muslim couples.

By examining factors like financial stability, religious values, and the quality of interpersonal communication, this study hopes to provide valuable insights for reducing divorce rates and improving marital satisfaction within the Muslim community. The outcomes are expected to support Islamic counseling centers, religious institutions, and national policymakers in designing more effective, evidence-based marriage support programs.

THE BACKGROUND

Marriage has long been considered the cornerstone of social harmony in Islamic communities, providing spiritual companionship and contributing to the overall balance of society. However, Malaysia is currently facing a troubling shift in this area, with declining marriage rates and a rise in divorce cases among Muslim couples. According to the Department of Statistics Malaysia (2023), the number of Muslim marriages dropped by 16.8% from 168,467 in 2022 to 140,176 in 2023, while Muslim divorces accounted for 76.6% of all divorce cases, reflecting a growing concern that needs to be addressed.

Most existing studies on marital success have relied heavily on qualitative approaches, such as interviews, personal stories, and expert opinions. While valuable, these methods often lack the statistical depth and objectivity that data-driven analysis can offer. This research aims to bridge that gap by introducing quantitative methods into the study of marital success, applying descriptive, predictive, and prescriptive analytics through tools like statistical correlation, machine learning classification models, and sentiment analysis.

Using a combination of primary survey data from Malaysian Muslim couples and secondary data from sources like Kaggle, the study will identify patterns and key factors that influence marital stability. Beyond contributing new knowledge to existing literature, the findings will showcase how data science can be practically applied to tackle meaningful social issues and inform better decision-making for communities and policymakers alike.

THE PROBLEM STATEMENT

A variety of interconnected factors, such as financial security, religious commitment, communication quality, and family influences, play an important role in shaping the success of a marriage. Yet, many Muslim couples in Malaysia lack access to objective, data-driven insights about what truly contributes to a long-lasting and stable marriage.

Recent national statistics highlight this concern, showing a 16.8% drop in Muslim marriage rates and that over three-quarters of all divorce cases in Malaysia involve Muslim couples. Despite these worrying trends, many Islamic counseling services and family advisory organizations continue to rely mainly on anecdotal evidence and personal experiences rather than empirical, data-backed analysis to guide their advice and interventions.

This study aims to fill that gap by identifying and examining the key factors that influence the success of Muslim marriages using data analytics and machine learning techniques. The lack of established, quantitative models predicting marital success within Muslim communities represents a notable gap in current research. Through this project, the goal is to deliver actionable, data-informed insights that can help couples, marriage counselors, and policymakers develop more effective strategies to strengthen marriages and reduce divorce rates in the Muslim community.

THE RESEARCH QUESTION AND HYPOTHESIS

Research Question

We have identified several key questions to investigate in this study about the elements that influence Muslim marriage success. The following research questions are going to guide our analysis:

1. How do financial stability and employment status affect the probability of a successful marriage in Muslim couples?

2. What role does religious connection have in guaranteeing long-term satisfaction in marriage and lowering the probability of divorce?

3. In what ways does good communication between spouses influence marriage success, and how does it relate to other variables like financial or cultural expectations?

Research Hypothesis

To test the factors influencing the success of Muslim marriages in Malaysia, we propose the following hypotheses:

1. Hypothesis 1: Financial stability and employment status have a positive correlation with marital success, as economic security reduces stress and conflict between spouses.
2. Hypothesis 2: Religious commitment and shared Islamic values between spouses significantly contribute to long-term marital satisfaction and reduce the likelihood of divorce.
3. Hypothesis 3: Effective communication between spouses mediates the relationship between financial stability and marital success, enhancing conflict resolution and overall relationship satisfaction.

THE RESEARCH OBJECTIVES

The purpose of this research is to:

1. Analyze the impact of financial stability and employment status on Muslim marriage success by identifying patterns of economic security that contribute to lower divorce rates.
2. Examine the role of religious commitment in fostering long-term marital satisfaction by assessing how shared Islamic values influence relationship stability.
3. Investigate the effect of communication quality on marital success and explore its interaction with financial and cultural expectations in maintaining healthy relationships.
4. Develop a predictive model using data analytics techniques (statistical correlation, machine learning, and sentiment analysis) to determine the most significant factors contributing to successful Muslim marriages.

5. Provide data-driven insights for marital counselors, religious institutions, and policymakers to develop strategies that promote marriage stability and reduce divorce rates in Malaysia.

THE RESEARCH SIGNIFICANCES

This study is important because it highlights how government organisations, lawmakers, religious leaders, and marriage counsellors can help address the increasing divorce rate. This study gives stakeholders important insights into early warning signals of divorce by applying machine learning to analyse relationship patterns and personal aspects. This allows for targeted treatment and assistance. The results may be used to increase the efficiency of initiatives that support marriage and foster family stability at organisations like the Ministry of Women, Family and Community Development, and religious organisations. Professionals have been provided with a trustworthy tool to evaluate relationship potential risks before situations become permanent, thanks to the dataset-driven methodology.

Understanding the main signs of divorce gives everyone the ability to recognise problematic marital patterns and seek counselling sooner rather than later. This knowledge may reduce the chance of divorce or separation by encouraging better communication, understanding, and more rational choices in marriages.

This study can also help marriage counsellors and religious advisers by providing an in-depth understanding of common risk factors leading to divorce. This allows them to properly manage time and resources, offer targeted advice, and assist couples in treatment sessions by utilising evidence-based techniques.

The results can be used by social development officers and policymakers to improve premarital education programs, create community-based psychoeducational programs, and create future laws that support families. In the end, by integrating technology with interpersonal interactions and wellbeing, this study helps to create a society that is stronger and aware.

THE RELEVANT WORKS / LITERATURE REVIEW

The use of artificial intelligence in medical diagnostics has grown significantly, with recent studies focusing on predictive modeling for heart disease. Madhumita Pal and Smita Parija (2021) utilized a dataset of 303 patient records to develop a Random Forest classification model capable of predicting heart disease presence. Using Python and 10-fold cross-validation, the model achieved an accuracy of 86.9%, with a sensitivity of 90.6% and specificity of 82.7%. Their analysis also produced an AUC score of 93.3%, indicating strong classification performance. This study highlights the potential of machine learning algorithms to support early detection and decision-making in cardiovascular healthcare.

Sharma, Chudhey & Singh (2021) developed a machine-learning framework to predict divorce cases using multiple classification algorithms. They tested six models: Perceptron, Decision Tree, Random Forest, Naïve Bayes, K‑Nearest Neighbors, and Support Vector Machine, along with Logistic Regression, Stochastic Gradient Descent, and Multilayer Perceptron architectures. Models were evaluated using various data partitions (50:50, 66:34, 80:20) and 10-fold cross-validation, on both full datasets and reduced feature sets (6 or 7 attributes).

Mian Muhammad Sadiq Fareed et al. (2022) proposed an ensemble learning framework combining Support Vector Machine (SVM), Passive Aggressive Classifier, and Multilayer Perceptron (neural network) to predict divorce prospects. They developed a questionnaire-based dataset of 54 marriage-related features, validated by field specialists, and applied 5-fold cross-validation. The combined model achieved 100 % accuracy, with ROC, recall, precision, and F1 scores all near 97 %, showcasing high reliability. Their analysis also identified key indicators most significant in divorce prediction, offering valuable insights for marital counseling and early intervention strategies

Moumen et al. (2024) examined divorce prediction in Saudi Arabia's Hail region using the 54‑item Divorce Predictor Scale (DPS), grounded in Gottman couples therapy principles. Surveying 148 individuals (116 married, 32 divorced), the study compared three machine learning models, Artificial Neural Network (ANN), Naïve Bayes (NB), and Random Forest (RF)—both before and after feature selection. Using correlation-based feature selection, they narrowed the dataset to the six most impactful DPS items. Results revealed that RF achieved the highest accuracy at 91.66%, outperforming NB (88.14%) and ANN (80%). The authors concluded that DPS is effective for early divorce prediction and can assist counselors in formulating interventions based on the Gottman framework.

Aimran et al. (2022) analyzed divorce prediction among Malaysian women using machine learning on data from 7,226 ever-married women aged 15–59, sourced from the Malaysia Population and Family Survey (MPFS-5). They compared six models: Decision Tree (C5.0, CHAID), Logistic Regression (forward/backward), and Artificial Neural Network (MLP, RBF). The Decision Tree with C5.0 achieved the highest accuracy at 77.96%, outperforming MLP (74.68%) and Logistic Regression (67.89%). Key predictors identified were wives' employment status, husbands' employment, type of marriage, race/ethnicity, long-distance relationships, wives' education level, age group, and religion.

Milani et al. (2020) investigated the effectiveness of emotional intelligence (EI) training on marital satisfaction among 60 married women in Iran. The study used a quasi-experimental pretest-posttest control group design, measuring marital satisfaction, sexual quality of life, and psychological well-being before and after the intervention. Results showed statistically significant improvements ($p < 0.01$) in all three variables in the experimental group. The findings suggest that EI training is an effective predictor and enhancer of marital success, particularly in emotional and sexual domains.

Biggiogera et al. (2021) analyzed communication behavior prediction during couples' conflict interactions using a dataset of 368 German-speaking Swiss couples engaged in 8-minute conflict discussions, segmented into 10-second intervals. They compared models using TF-IDF, LIWC, openSMILE, and BERT features, training an SVM with RBF kernel. The BERT-only model achieved the highest balanced accuracy at 69.4%, outperforming LIWC (65.4%), TF-IDF (65.6%), and openSMILE (61.3%), while combining BERT with paralinguistic features did not improve performance (69.2%). The study confirmed that contextual language embeddings (BERT) are superior to traditional LIWC for predicting nuanced communication behaviors in couples. They also noted that adding paralinguistic features (via openSMILE) does not significantly increase performance, suggesting a primary role for verbal semantics in conflict analysis.

Vowels et al. (2021) investigated predictors of sexual desire using a machine learning approach on combined samples of 1,846 individuals (754 forming 377 couples). They applied a Random Forest model with Shapley Additive Explanations to identify the most salient predictors of both dyadic and solitary sexual desire. The models explained approximately 40% of the variance in both types of desire. Key predictors for dyadic desire included sexual satisfaction and romantic love, while masturbation frequency and permissive sexual attitudes were most important for predicting solitary desire. Additionally, gender did not significantly predict desire levels, and partner effects (especially romantic love and sexual satisfaction) also contributed meaningfully. This study highlights that relationship satisfaction and emotional intimacy play central roles in sexual desire among couples, providing machine learning–based evidence to guide future interventions in couple therapy.

Azhar & Mohd Hoesni (2023) explored the link between emotional intelligence (EI) and marital satisfaction among Malaysian Muslim married individuals during the COVID-19 pandemic. Data were collected through an online survey from approximately 300 respondents, using the Schutte Self-Report Emotional Intelligence Test (SSEIT) to assess EI and the Kansas Marital Satisfaction Scale (KMSS) to measure marital satisfaction. Using Pearson correlation analysis, the study found a weak but statistically significant positive correlation between emotional intelligence and marital satisfaction, particularly in the emotional usage subscale ($r = 0.10$, $p < 0.05$). This indicates that while individuals with higher emotional intelligence tended to report slightly higher marital satisfaction, the strength of the relationship was minimal. The authors argue that EI remains an important factor in marital dynamics, especially during emotionally stressful periods such as the pandemic, but acknowledge that other contextual and interpersonal variables may exert a stronger influence. They recommend that future research adopt a mixed-method approach and a more diverse sample to capture a deeper understanding of how emotional skills interact with marital success in a Malaysian Muslim context.

Chen, Q., Zhang, M., Wu, W., Liu, D., Liu, T., & Yao, Y. (2024) investigated the prognostic impact of marital status on survival outcomes in patients diagnosed with pancreatic ductal adenocarcinoma (PDAC) using data from the SEER database (2004–2015). This retrospective cohort study analyzed 43,400 patients, stratifying them into marital status categories: married, single, divorced/separated, and widowed. Survival outcomes were assessed using Kaplan–Meier curves and Cox proportional hazard models. Results showed that married patients had significantly better overall survival and cancer-specific survival than unmarried patients, even after adjusting for clinical and demographic covariates. The study concluded that marital status is an independent predictor of survival in PDAC, possibly due to better emotional support, earlier diagnosis, and treatment compliance among married individuals. The authors suggest that social support interventions may be particularly beneficial for unmarried cancer patients.

Table 1: Relevant Works with the Project

No.	Year	Author	Research problem	Main technique	Result	Future works
1	2021	Madhumita Pal and Smita Parija	Predict heart disease	Random Forest	90.6% of patients with heart disease were correctly classified.82.7% of patients without heart disease were correctly classified Correctly predicts the patient using random forest	Use claude computing to manage high volume of patient data
2	2021	Aditya Sharma,Arshdeep Singh Chudhey, Mrityunjay Singh	The increasing number of divorces all over the world.	Perceptron classifier, Decision Tree classifier, Random Forest classifier, Naive Bayes classifier, K-Nearest Neighbour classifier and Support Vector Machine classifier	Get highest score accuracy (98.5%) using the Perceptron model.	Use Feature selection in the future which can help to decrease training time and increase the accuracy of the model.
3	2022	Mian Muhammad Sadiq Fareed et al.	Main factor of divorce cases happens.	Support Vector Machine, Linear Model, and Neural Network	Ensemble learning (EL) achieved the highest accuracy of 100%.	Enhance the questionnaire dataset. Apply the data augmentation techniques. Explore different deep learning models.
4	2024	Abdelkader Moumen, Ayesha Shafqat, Tariq Alraqad, Etaf Saleh Alshawarbeh, Hicham Saber & Ramsha Shafqat	Divorce Prediction in KSA Using ML	Artificial Neural Network, Naive Bayes , and Random Forest	Accuracy for : ANN is 80.00% RF is 88.14% NB and is 91.66%,	Examine the effectiveness of the Gottman couples therapy model's intervention strategies in the Hail region
5	2022	Aimran, N., Rambli, A., Afthanorhan, A., Mahmud,	Predict Malaysian women divorced	Decision Tree, Logistic Regression and Artificial Neural	Decision tree obtain the highest accuracy (77.96%) compare to LR and	-

		A., Sapri, A., & Aireen, A	using Machine Learning	Network	ANN	
6	2022	Milani, A. S., Hosseini, M., Matbouei, F., & Nasiri, M	Effect of emotional intelligence training on marital satisfaction	Quasi-experimental pretest-posttest with control group	Statistically significant improvement ($p < 0.01$) in marital satisfaction, sexual quality of life, and psychological well-being	-
7	2021	Biggiogera, J., Paleari, F. G., Regalia, C., & Fincham, F. D	Predicting communication behavior in couples' conflict	SVM with TF-IDF, LIWC, openSMILE, and BERT	BERT-only model achieved highest balanced accuracy (69.4%)	Extend to real-time conflict intervention applications using ML
8	2021	Vowels, L. M., Mark, K. P., & Leonard, L. M.	Predictors of sexual desire in couples	Random Forest + SHAP analysis	Explained ~40% variance in desire; key predictors: love, satisfaction, permissive attitudes	Expand predictive models for sexual satisfaction using larger ML datasets
9	2023	Azhar, N. I., & Mohd Hoesni, S.	Emotional intelligence and marital satisfaction during COVID-19	SSEIT & KMSS with Pearson correlation	Weak but significant positive correlation (r = 0.10, $p < 0.05$)	-
10	2024	Chen, Q., Zhang, M., Wu, W., Liu, D., Liu, T., & Yao, Y.	Impact of marital status on pancreatic cancer survival	Kaplan–Meier & Cox regression analysis	Married patients had significantly better overall and cancer-specific survival	-

METHODOLOGY

This section describes the methodology process, which used the primary and secondary datasets to identify significant factors affecting Muslim marriage success using data analysis

1. **Primary Dataset:**

- **Name and Source of the Dataset:**
    - The dataset was gathered using a Google Form survey entitled "**Survey: Factors Affecting Muslim Marriage Success in Malaysia**".
    - A total of 50 replies were collected from married Muslim persons in Malaysia.
    - The dataset includes 22 features obtained from six sections of the survey, including:
        - Financial Stability
        - Communication Quality
        - Religious Commitment
        - External Influences
        - Emotional and marital satisfaction indicators
    - The final column (target variable) is Marriage Satisfaction, represented as:
        - 0 = Not stable
        - 1 = Stable
        - 0.5 = Maybe (Responses have been removed from model training to ensure binary classification.)

- **Tool Used:**
    - Platform: **Google Colab**
    - Language: **Python**
    - Libraries:
        - **Pandas** and **Numpy** are used for data preprocessing
        - **Matplotlib** and **Seaborn** are used for data visualization
        - **Scikit-learn** is used for machine learning and model evaluation

- **Machine Learning and Algorithm Used:**
  - We utilized Random Forest Classifier, which is an ensemble learning algorithm that creates numerous decision trees and aggregates the findings using majority voting. This approach is to increase the prediction accuracy and prevent overfitting.

- **Training and Testing Dataset:**
  - The dataset was initially filtered to exclude responses with "Maybe" values (0.5) from the target column.
  - The dataset was then split in an 80/20 ratio between training and testing:
    - Training set: 80% of valid responses.
    - Testing set contains 20% of valid responses.
  - The split was done with Scikit-learn's train_test_split() and random_state = 42.
  - In addition, stratified 5-fold cross-validation was utilized to ensure robust evaluation and eliminate bias induced by the small sample size.

- **Performance Measured Used:**
  - **Accuracy**: Measures the overall correctness of predictions.
  - **F1-Score: Balances precision and recall, which is especially beneficial in tiny, imbalanced datasets.**
  - **ROC Curve and AUC Score**: Measure the ability to differentiate between classes.
  - **Classification Report and Confusion Matrix:** Provide a comprehensive overview of true positives, false positives, recall, and precision.

- **Statistical Test Used:**
  - **Correlation Analysis**: A Pearson correlation heatmap was used to determine linear correlations between numeric features and the desired outcome (marriage stability). This helped identify which survey responses (e.g., EmotionalWellbeing, MarriageSatisfaction, SameIslamicValues) had the strongest link with perceived marriage stability.

  - There was no mean comparison because of the minimal class diversity after removing "Maybe" responses.

- **Random Forest Diagram:**

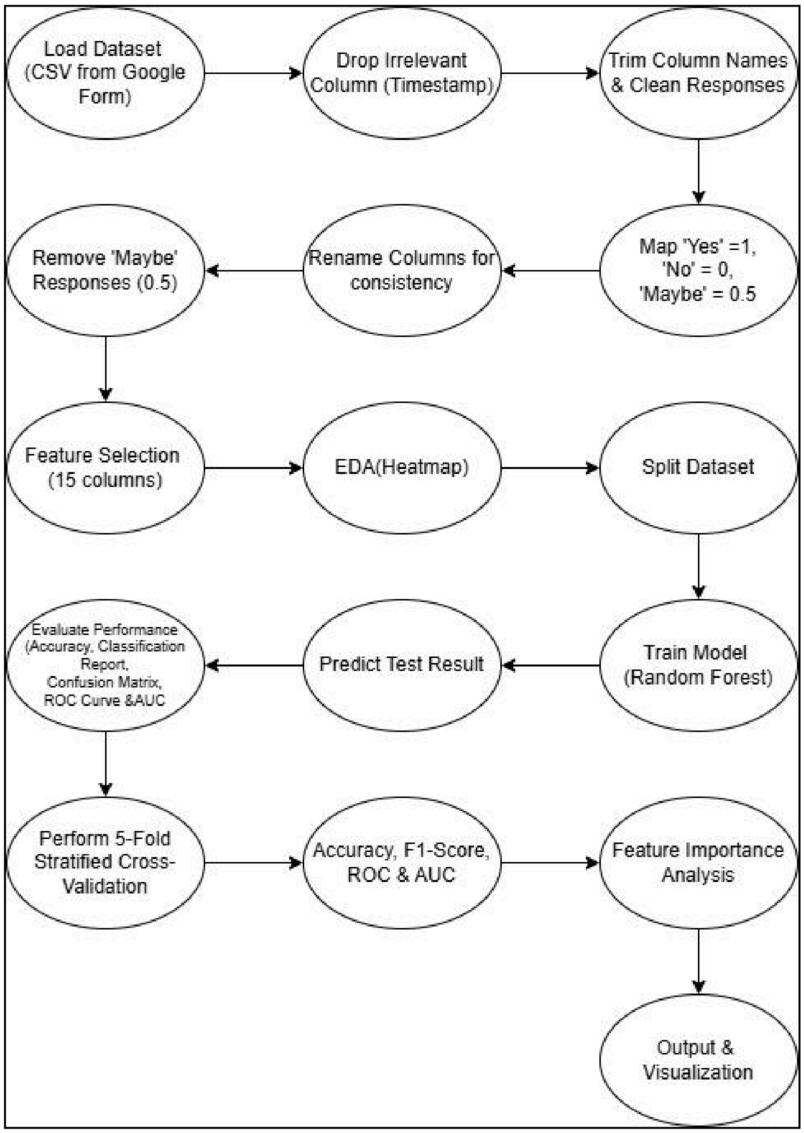

Figure 1: Random Forest Diagram (Primary Dataset)

2. **Secondary dataset:**

- **Name and Source of the Dataset:**
  - The dataset used for the secondary dataset is the "Divorce Predictors" dataset, which is available publicly on Kaggle.
  - There are 170 responses and 55 columns, with the first 54 columns (Q1 to Q54) representing responses to questions from the Divorce Predictors Scale (DPS), which is based on Gottman couples therapy.
  - The responses are evaluated on a 5-point scale:
    - 0 = Never
    - 1 = Seldom
    - 2 = Averagely
    - 3 = Frequently
    - 4 = Always
  - The final column (target variable) is Divorce, represented as:
    - 0 = Not Divorce
    - 1 = Divorce

- **Tool Used:**
  - Platform: **Google Colab**
  - Language: **Python**
  - Libraries:
    - **Pandas** and **Numpy** are used for data preprocessing
    - **Matplotlib** and **Seaborn** are used for data visualization
    - **Scikit-learn** is used for machine learning and model evaluation

- **Machine Learning and Algorithm Used:**
  - We utilized Random Forest Classifier, which is an ensemble learning algorithm that creates numerous decision trees and aggregates the findings using majority voting. This approach is to increase the prediction accuracy and prevent overfitting.

- Random Forest Classifier performs well with tabular datasets, including mixed or categorical values, manages feature importance estimation, and is Robust against overfitting on smaller datasets.

- **Training and Testing Dataset:**
    - The dataset was split at an 80/20 ratio:
        - Training set: 136 samples
        - Testing set: 34 samples
    - The split was performed using Scikit-learn’s train_test_split() with a random_state = 42.
    - Additionally, stratified 5-fold cross validation was used to evaluate the model's generalizability.

- **Performance Measured Used:**
    - **Accuracy**: Measures the overall correctness of predictions.
    - **Confusion Matrix**: Distribution of true and false positives, etc.
    - **Classification Report**: Includes precision, recall, and F1 score for both classes.
    - **ROC Curve and AUC Score**: Measure the ability to differentiate between classes.

- **Statistical Test Used:**
    - **Correlation Analysis**: To investigate the linear relationship between item scores and divorce likelihood, Pearson correlation was used for each question (Q1 – Q54) and the target variable, Divorce. This allowed for determining which specific questions were most strongly connected with the target result.

    - **Mean Score Comparison**: The dataset was divided by Divorce label (0 = Not Divorced, 1 = Divorced), and the average score for each item was given. This allowed a quantitative comparison between how divorced and non-divorced respondents ranked each question.

- **Random Forest Diagram:**

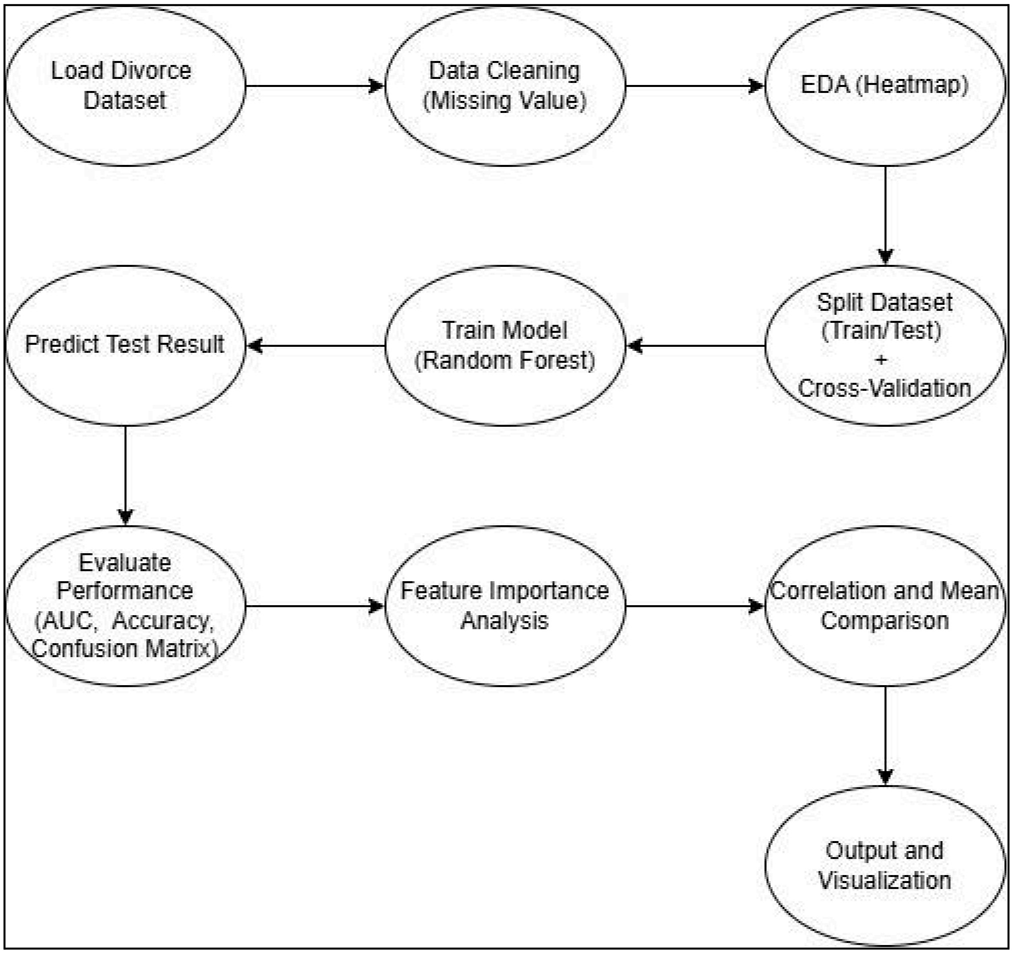

Figure 2: Random Forest Diagram (Secondary Dataset)

RESULTS

This section summarizes the findings from both the primary datasets (Google Form survey) and the secondary dataset (Kaggle). The results are presented in tables and supported by relevant visuals.

1. **Primary Dataset**

- **Summary of Descriptive Statistics:**

Table 2: Summary of Descriptive Statistics from Primary Dataset

Feature	Most Frequent Category	Count
Age Range	26-35	16
Education Level	Degree	25
Employment Status	Employed	28
Monthly Income	RM 3001 - RM 5000	18
Years Married	1-3 Years	16

- **Visualization of Key Demographic Distributions:**

Code :

```
# Age range
sns.countplot(data=df, x='AgeRange',
order=df['AgeRange'].value_counts().index)
plt.title('Distribution of Age Range')
plt.xticks(rotation=45)
plt.show()
```

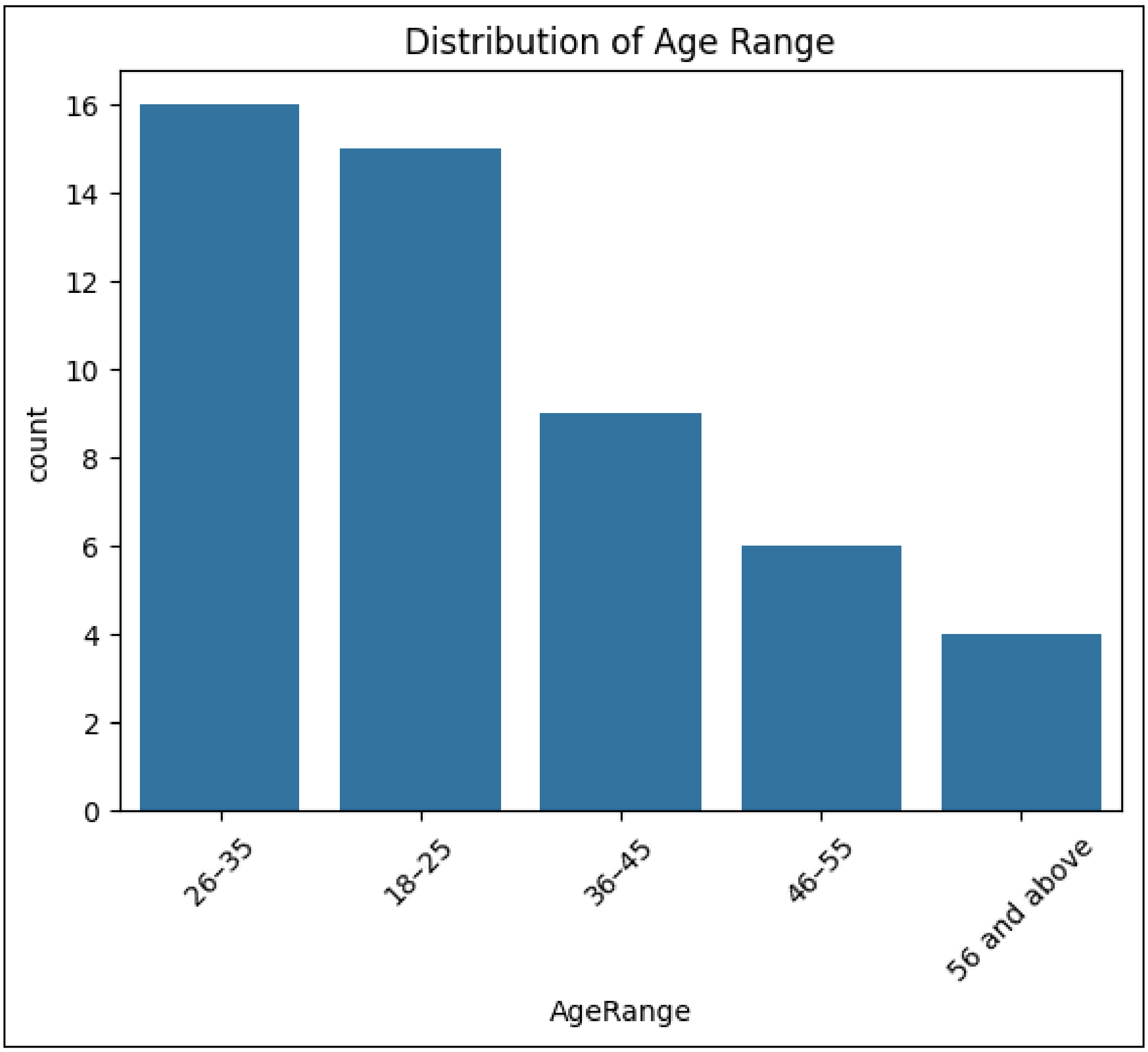

Figure 3: Age Range

Code :

```
# Education
sns.countplot(data=df, x='Education',
order=df['Education'].value_counts().index)
plt.title('Distribution of Education')
plt.xticks(rotation=45)
plt.show()
```

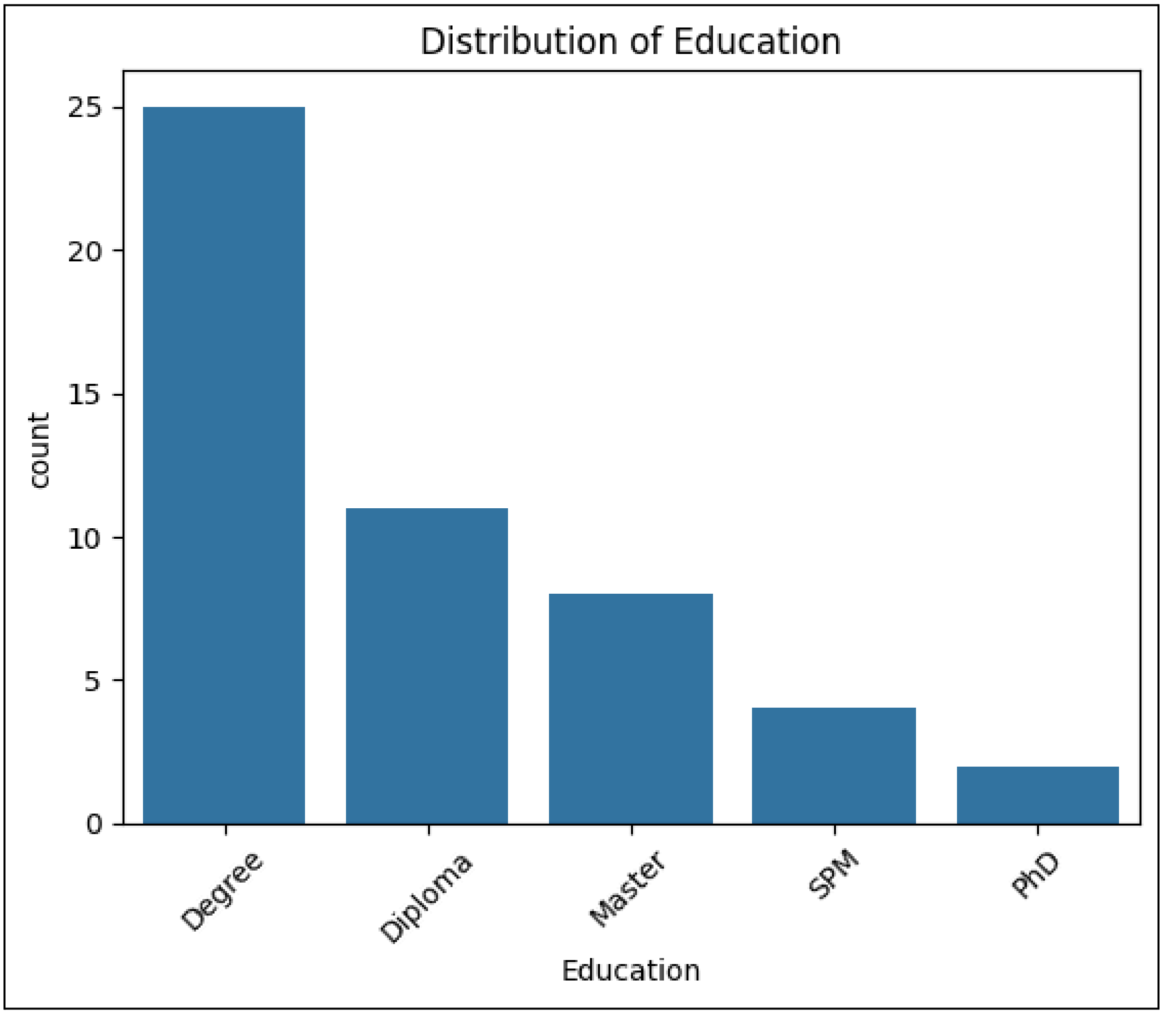

Figure 4: Education

Code :

```
# Employment
sns.countplot(data=df, x='Employment',
order=df['Employment'].value_counts().index)
plt.title('Distribution of Employment')
plt.xticks(rotation=45)
plt.show()
```

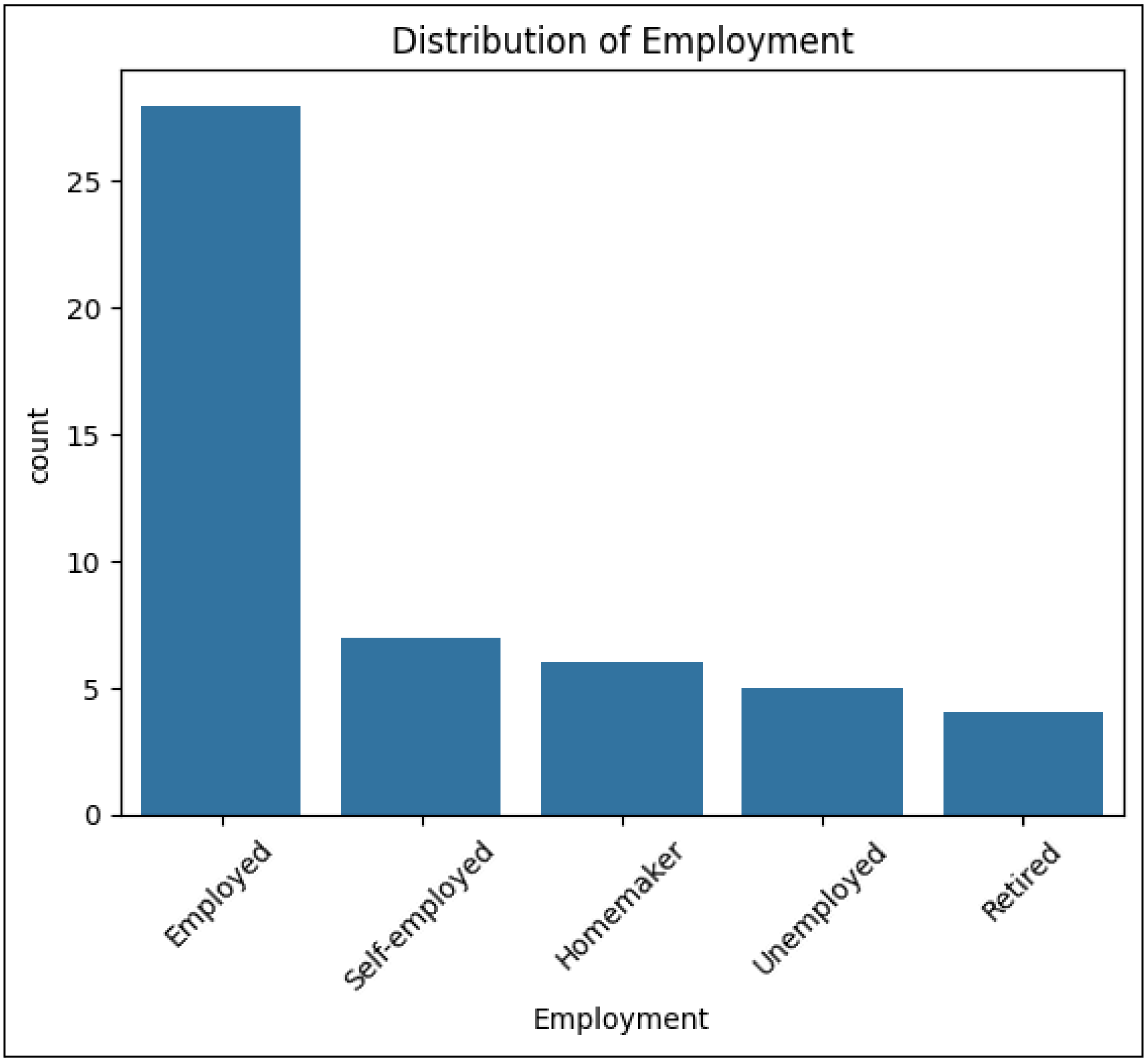

Figure 5: Employment

Code :

```
# Income
sns.countplot(data=df,
x='Income',order=df['Income'].value_counts().index)
plt.title('Distribution of Income')
plt.xticks(rotation=45)
plt.show()
```

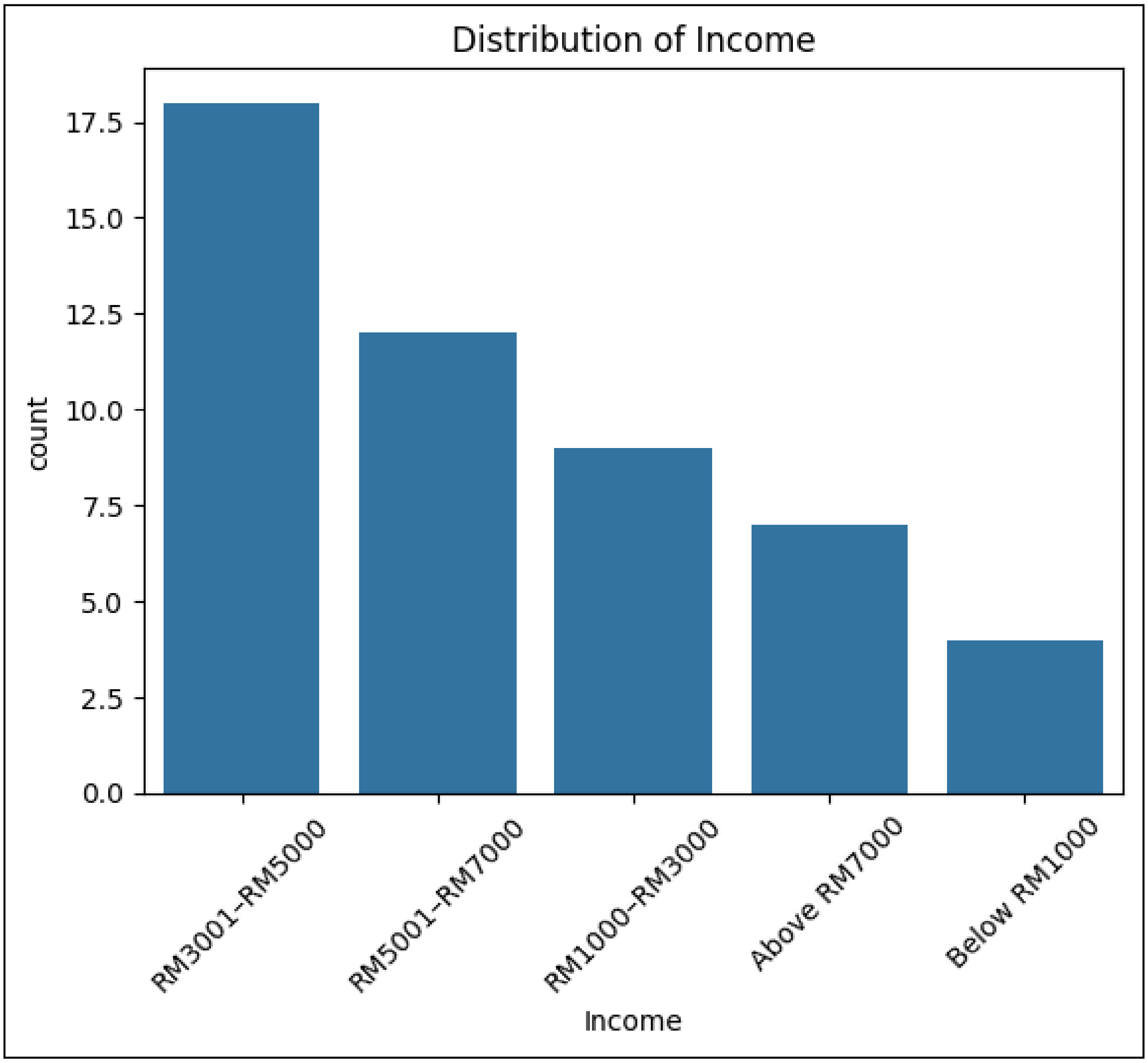

Figure 6: Monthly Income

Code :

```
# YearsMarried

sns.countplot(data=df, x='YearsMarried',
order=df['YearsMarried'].value_counts().index)
plt.title('Distribution of Years Married')
plt.xticks(rotation=45)
plt.show()
```

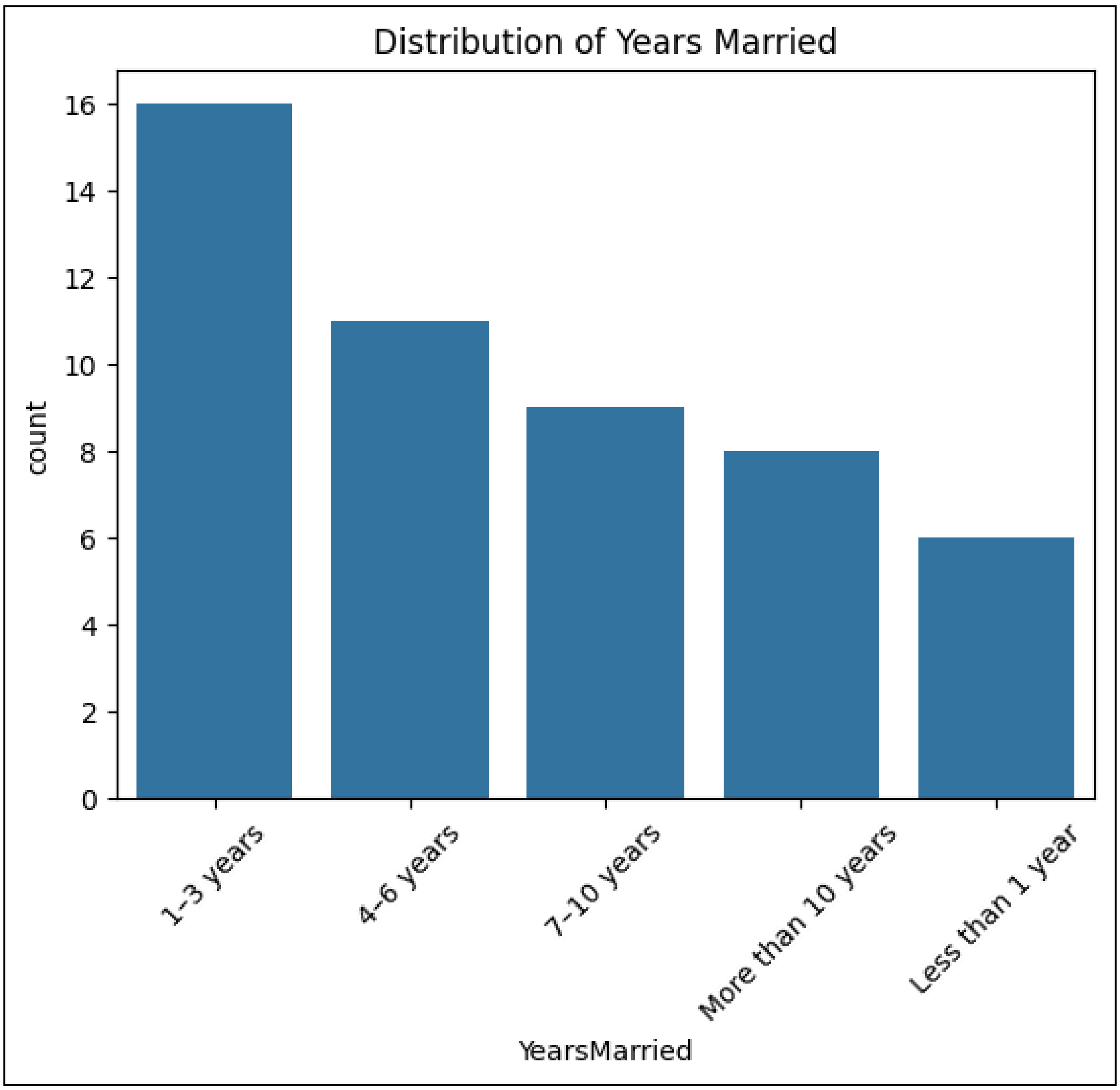

Figure 7: Years Married

- **Correlation Heatmap (Numeric Features Only):**

  Code :

```
numeric_df = df.select_dtypes(include='number')

plt.figure(figsize=(12, 10))
sns.heatmap(numeric_df.corr(), annot=True, cmap='coolwarm',
linewidths=0.5)
plt.title("Correlation Heatmap (Numeric Features Only)")
plt.show()
```

Table 3: Correlation Between Numeric Features and Marriage Stability

Feature	**Correlation with Marriage Stability**
ReligionCentrality	**0.91**
MarriageSatisfaction	**0.90**
FamilySupport	**0.87**
FinnancePlanning	**0.85**
StableIncome	**0.82**
SameIslamicValues	**0.82**

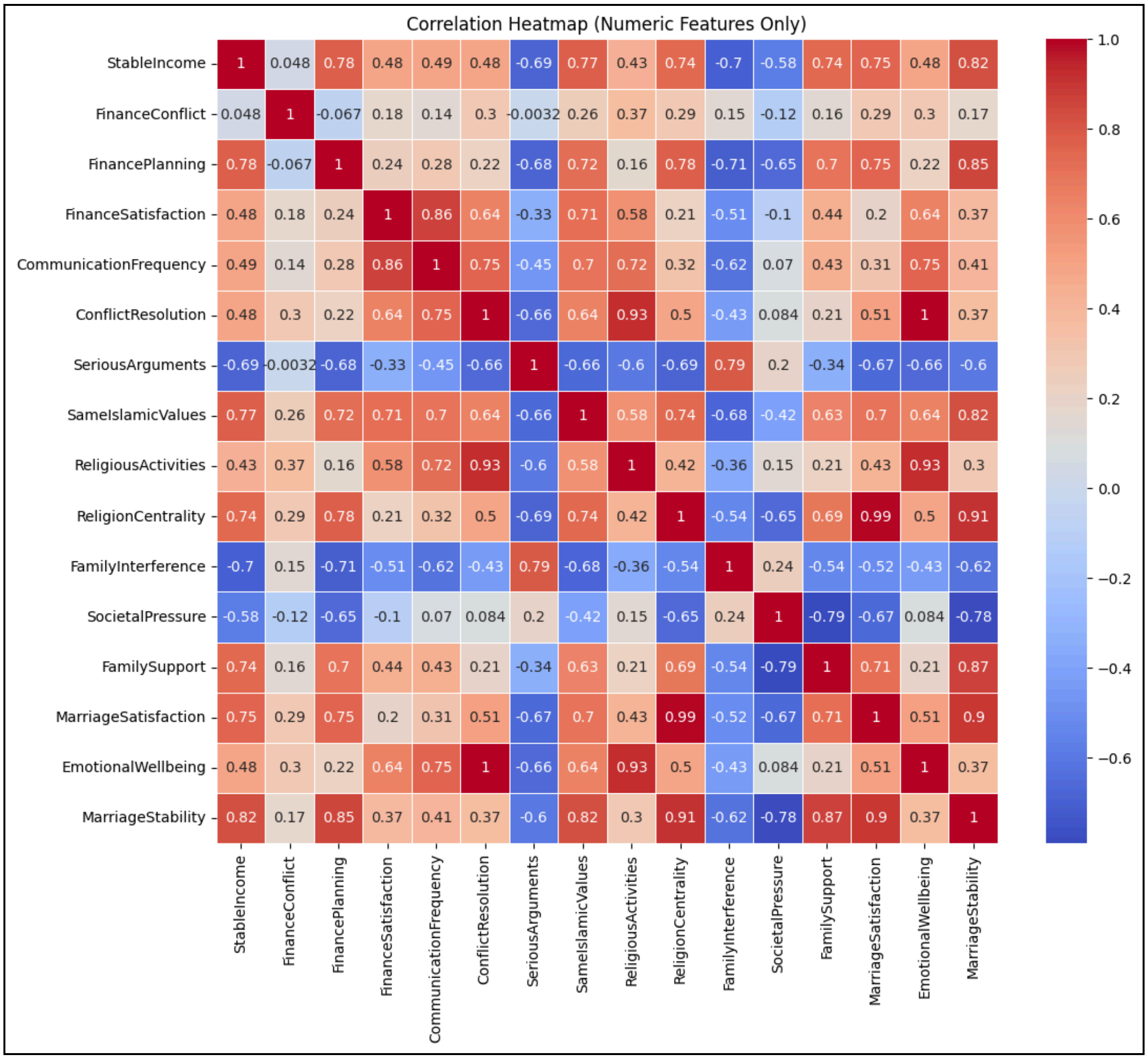

Figure 8 Correlation Heatmap (Numeric Features Only)

- **Random Forest Model Performance:**
    - After filtering out "Maybe" responses, the Random Forest Classifier was trained to predict marriage stability.

Code :

```
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestClassifier
from sklearn.metrics import accuracy_score, confusion_matrix,
classification_report, roc_curve, roc_auc_score

# To remove the "maybe" answer in target variables
df = df[df['MarriageStability']!=0.5]

target = 'MarriageStability'
features = [
    'StableIncome', 'FinanceConflict', 'FinancePlanning',
'FinanceSatisfaction',
    'CommunicationFrequency', 'ConflictResolution',
'SeriousArguments',
    'SameIslamicValues', 'ReligiousActivities',
'ReligionCentrality',
    'FamilyInterference', 'SocietalPressure', 'FamilySupport',
    'MarriageSatisfaction', 'EmotionalWellbeing'
]

X = df[features]
y = df[target]

X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2, random_state=42)
model = RandomForestClassifier(n_estimators=100, random_state=42)
model.fit(X_train, y_train)

y_pred = model.predict(X_test)
y_proba = model.predict_proba(X_test)[:, 1]
print("Accuracy:", accuracy_score(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test,
y_pred))
```

Table 4: Random Forest Model Performance on Primary Dataset

Metric	Value
Accuracy	1.00
Precision	1.00
Recall	1.00
F1-Score	1.00
AUC Score	1.00

Code :

```
cm = confusion_matrix(y_test, y_pred)
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues')
plt.title('Confusion Matrix')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.show()
```

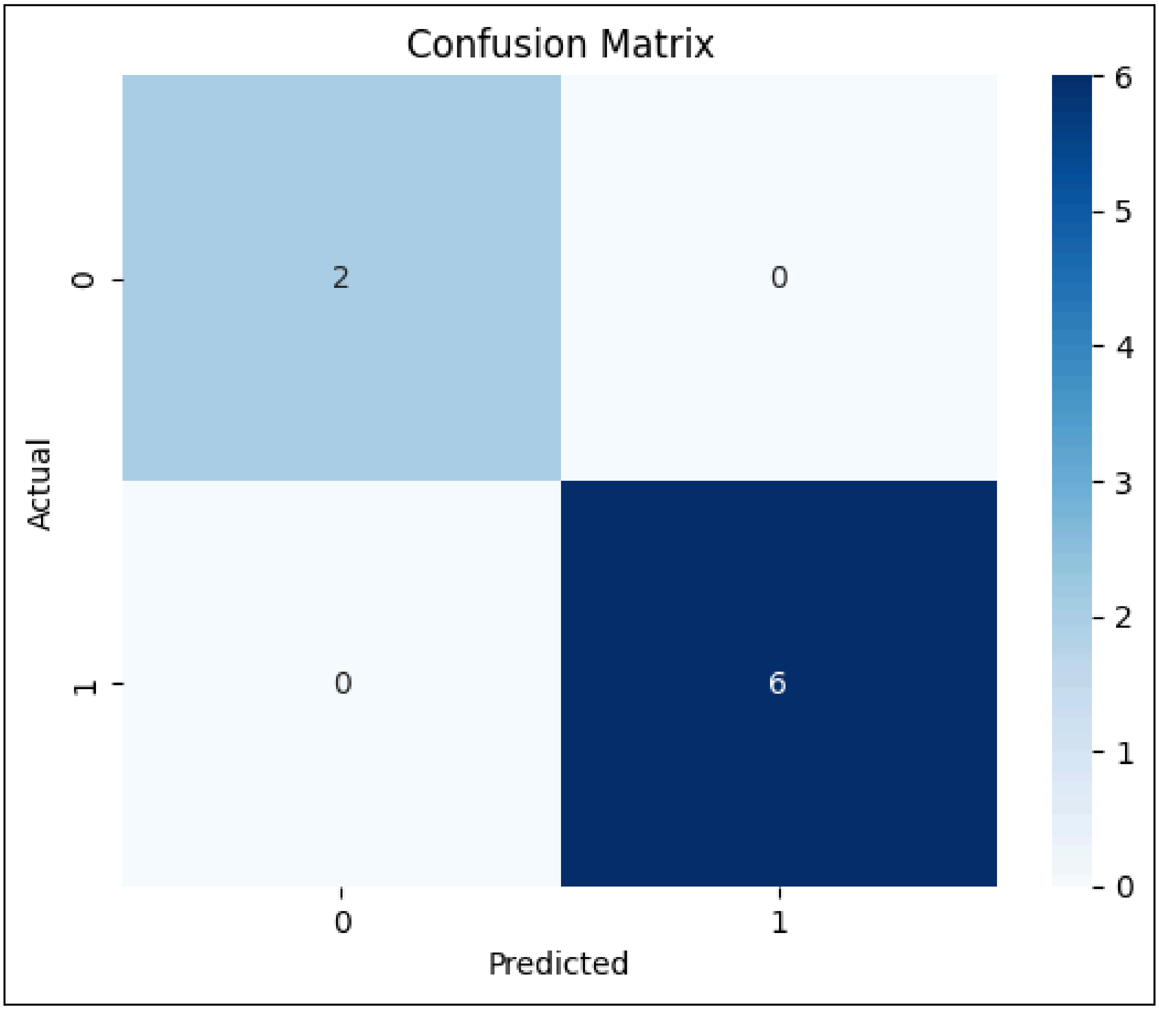

Figure 9: Confusion Matrix shows perfect classification results

Code :

```
fpr, tpr, _ = roc_curve(y_test, y_proba)
plt.figure()
plt.plot(fpr, tpr, label=f"AUC = {roc_auc_score(y_test,
y_proba):.2f}")
plt.plot([0, 1], [0, 1], linestyle='--')
plt.xlabel("False Positive Rate")
plt.ylabel("True Positive Rate")
plt.title("ROC Curve")
plt.legend()
plt.grid()
plt.show()
```

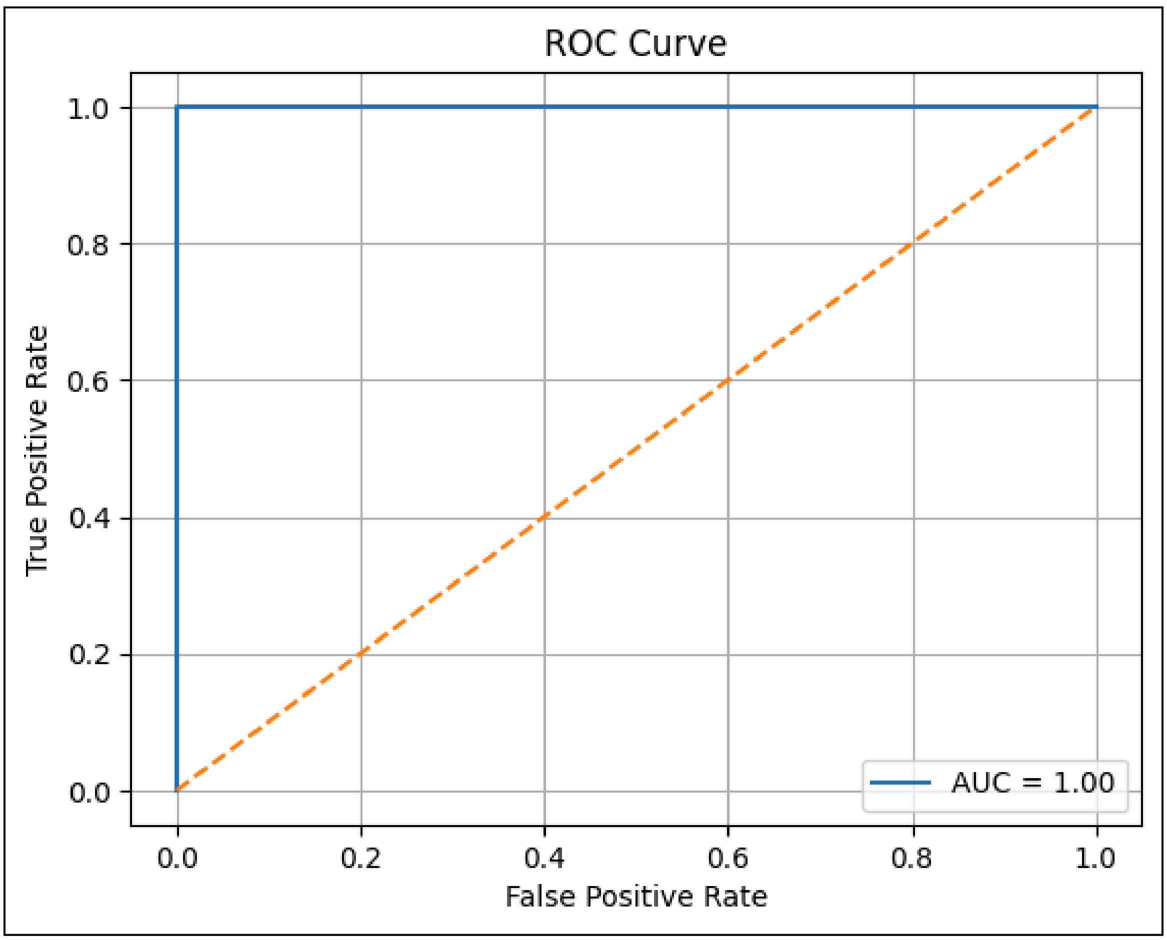

Figure 10: ROC Curve shows perfect classification results

- **Feature Importance (Top 10):**

  Code :

```
# Feature Importance
import pandas as pd

importances = pd.Series(model.feature_importances_, index=X.columns)
importances.nlargest(10).plot(kind='barh')
plt.title('Top 10 Important Features Predicting Marriage Stability')
plt.xlabel('Importance Score')
plt.ylabel('Feature')
plt.grid()
plt.show()
```

Table 5: Top 10 Important Features Predicting Marriage Stability

Feature	Importance Score
CommunicationFrequency	0.10
StableIncome	0.09
SeriousArguments	0.09
ReligionCentrality	0.09
FamilyInterference	0.09
MarriageSatisfaction	0.09
EmotionalWellbeing	0.09
FinnancePlanning	0.07
FinnanceSatisfaction	0.07
ConflictResolution	0.06

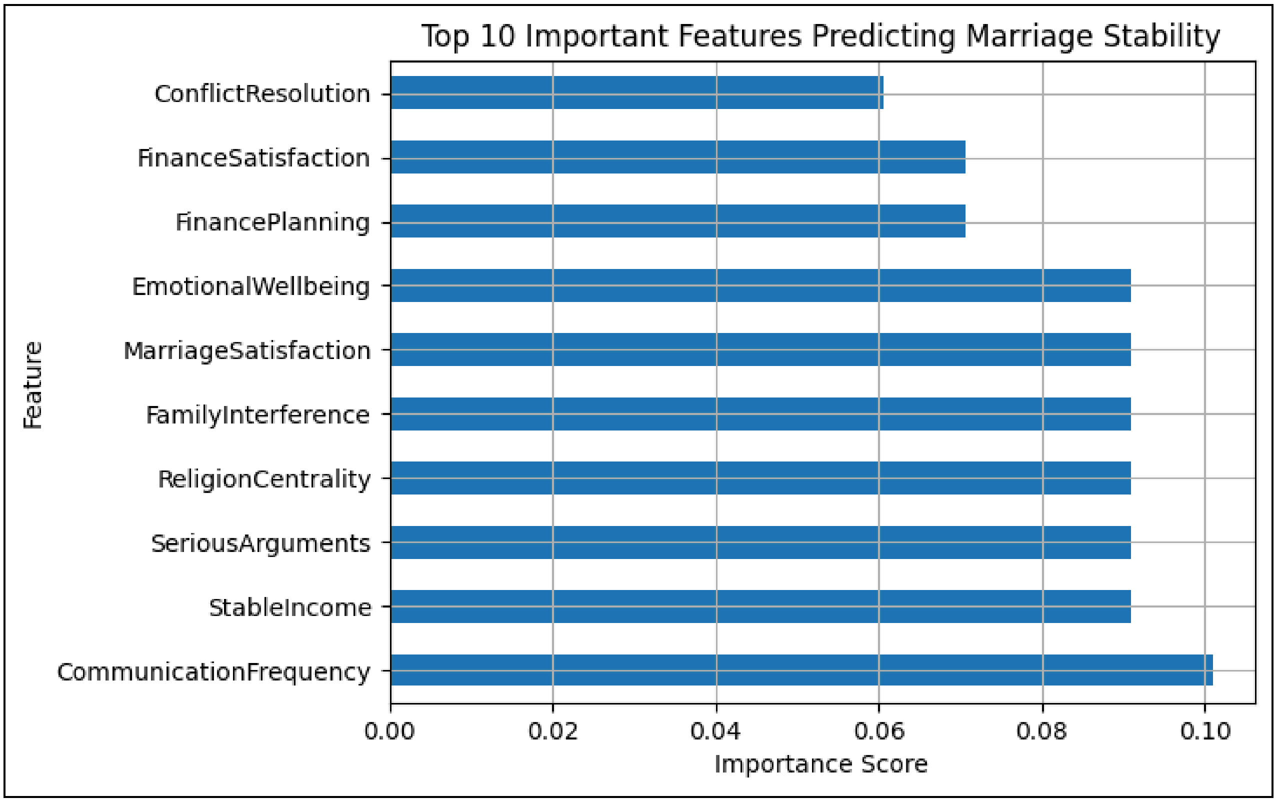

Figure 11: Top 10 Important Features Predicting Marriage Stability

2. **Secondary Dataset**

- **Class Distribution**

Code :

```
sns.countplot(x='Divorce', data=df)
plt.title('Divorce Distribution')
plt.xlabel('Divorce (1 = Yes, 0 = No)')
plt.ylabel('Count')

# Annotate bar values
for p in plt.gca().patches:
    plt.gca().annotate(f'{p.get_height()}',
                        (p.get_x() + p.get_width() / 2.,
p.get_height()),
                        ha='center', va='center', fontsize=11,
color='black', xytext=(0, 10),
                        textcoords='offset points')
plt.show()

print(df['Divorce'].value_counts())
print(df['Divorce'].value_counts(normalize=True) * 100)
```

Table 6: Class Distribution of Divorce Status

Class	Count	Percentage
Not Divorced (0)	84	49.4%
Divorced (1)	86	50.6%

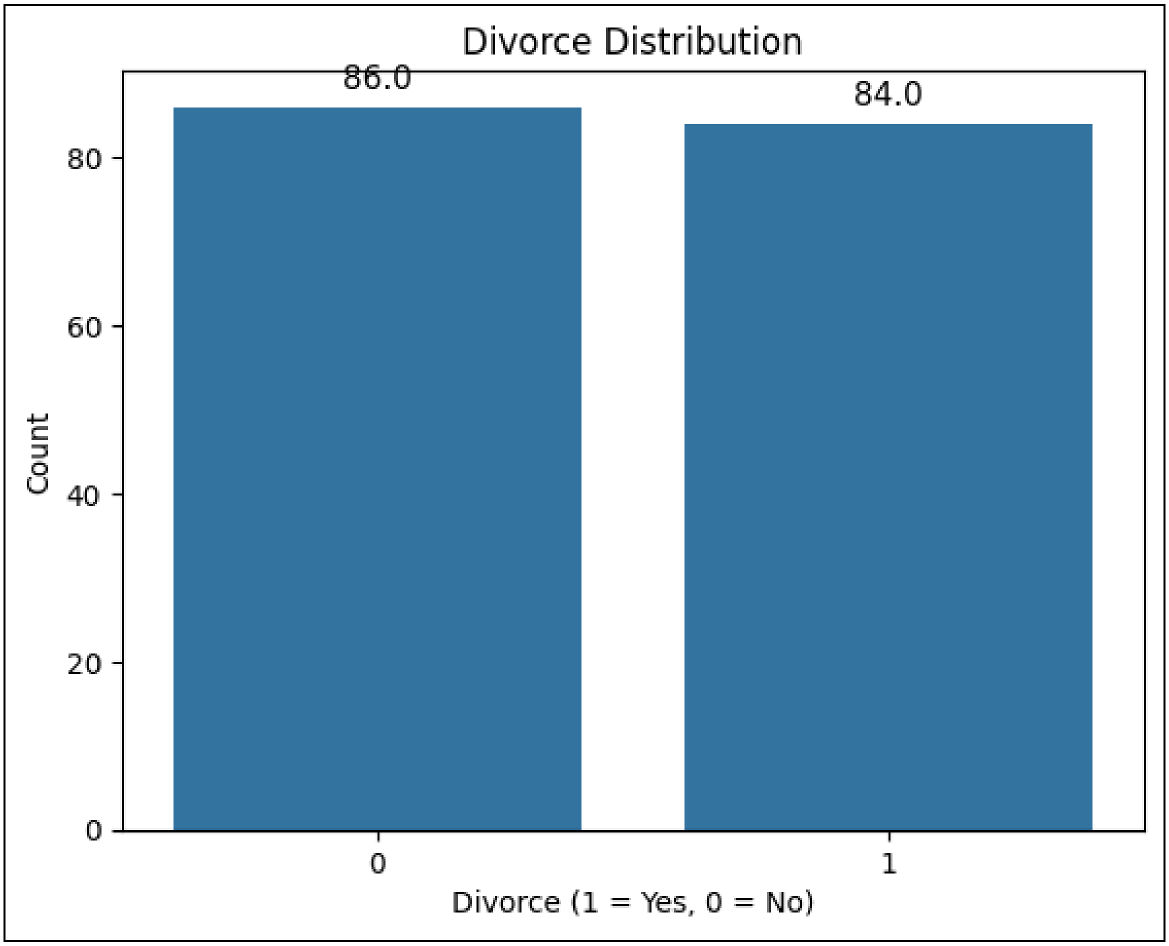

Figure 12: Distribution of Divorce Status (Balanced Classes)

- **Model Evaluation Result:**

  Code :

```
print("Accuracy:", accuracy_score(y_test, y_pred))
print("\nConfusion Matrix:\n", confusion_matrix(y_test, y_pred))
print("\nClassification Report:\n", classification_report(y_test,
y_pred))
```

Table 7: Model Evaluation Metrics for Secondary Dataset

Metric	Value
Accuracy	0.97
Precision	1.00
Recall	0.95
F1-Score	0.97
AUC Score	1.00

- **Confusion Matrix:**

  Code :

```
cm = confusion_matrix(y_test, y_pred)
sns.heatmap(cm, annot=True, fmt='d', cmap='Blues')
plt.title('Confusion Matrix')
plt.xlabel('Predicted')
plt.ylabel('Actual')
plt.show()
```

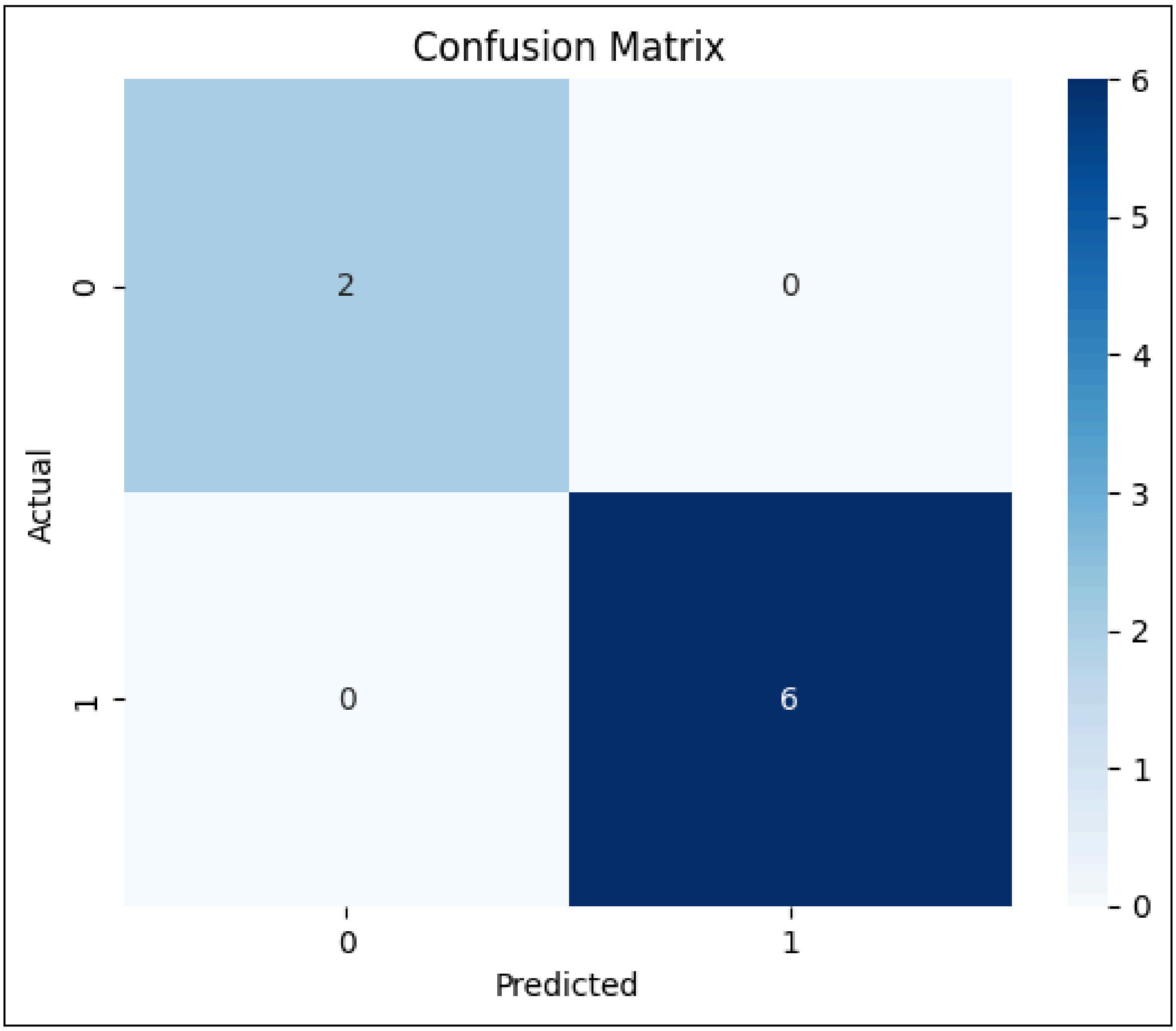

Figure 13: Confusion Matrix

- **ROC Curve:**
  - The model had an AUC score of 1.00, showing almost perfect classification abilities between divorced and non-divorced individuals.

    Code :

```
fpr, tpr, _ = roc_curve(y_test, y_proba)
plt.figure(figsize=(8,6))
plt.plot(fpr, tpr, label="Random Forest")
plt.plot([0, 1], [0, 1], 'k--')
plt.title("ROC Curve")
plt.xlabel("False Positive Rate")
plt.ylabel("True Positive Rate")
plt.legend()
plt.grid(True)
plt.show()

print("AUC Score:", roc_auc_score(y_test, y_proba))
```

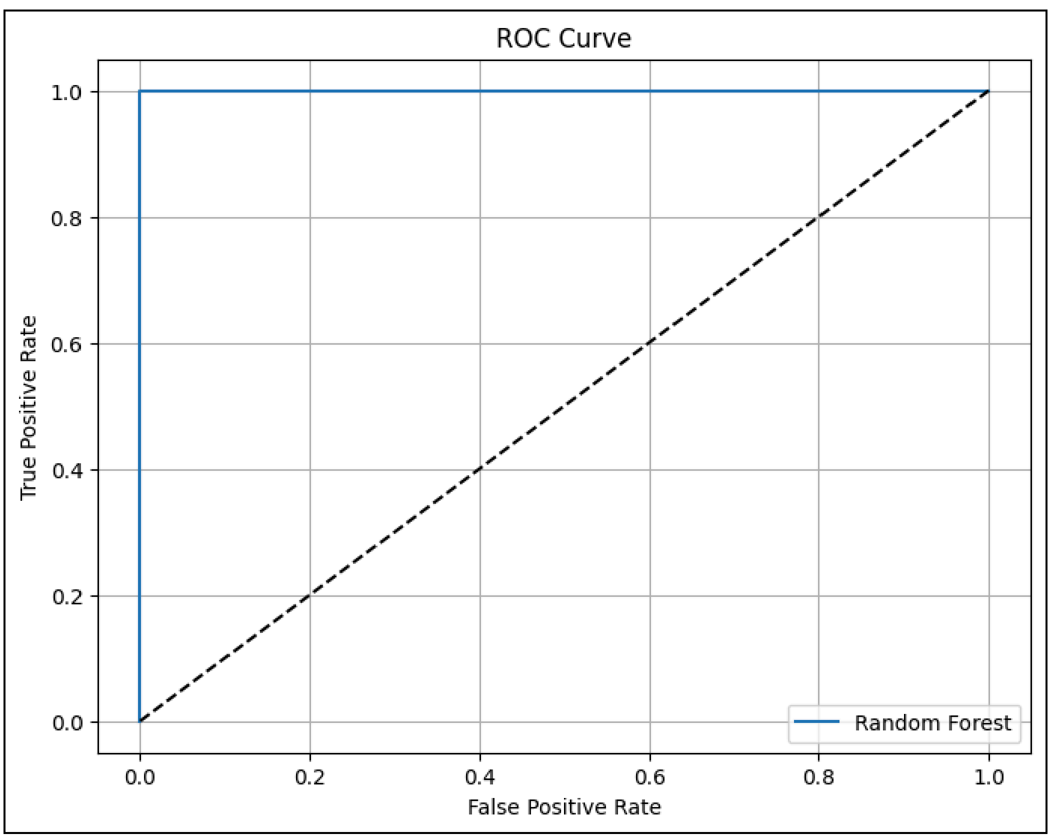

Figure 14: ROC Curve

- **Stratified 5-Fold Cross-Validation Results:**
  - **Mean Accuracy:** 0.976
  - **Std Deviation:** 0.022

Code :

```
cv = StratifiedKFold(n_splits=5, shuffle=True, random_state=42)
rf_cv = RandomForestClassifier(n_estimators=100, random_state=42)
cv_scores = cross_val_score(rf_cv, X, y, cv=cv, scoring='accuracy')

print("Cross-Validation Scores:", cv_scores)
print("Mean Accuracy:", cv_scores.mean())
print("Std Deviation:", cv_scores.std())
```

Table 8: Stratified 5-Fold Cross-Validation Accuracy Scores

Fold	Accuracy
1	0.97
2	1.00
3	0.94
4	0.97
5	1.00

- **Feature Importance (Top 10 Important Features):**

  Code :

```
feature_importance = pd.Series(model.feature_importances_,
index=X.columns)
top_features = feature_importance.nlargest(10)

plt.figure(figsize=(10,6))
top_features.plot(kind='barh')
plt.title("Top 10 Most Important Features in Predicting Divorce")
plt.xlabel("Feature Importance")
plt.gca().invert_yaxis()
plt.show()
```

Table 9: Top 10 Important Features for Predicting Divorce

Feature	Importance Score
Q40	0.098
Q17	0.096
Q18	0.095
Q19	0.093
Q12	0.091
Q20	0.067
Q16	0.058
Q11	0.057
Q15	0.049
Q26	0.044

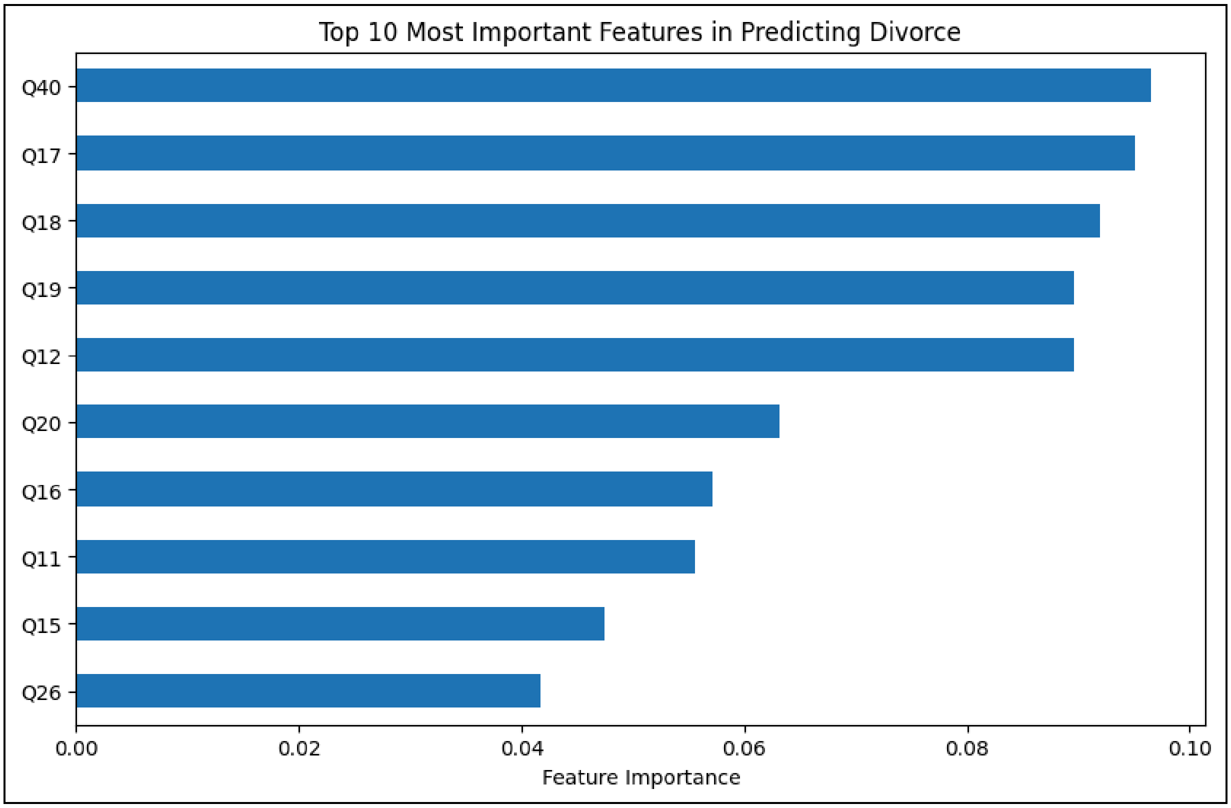

Figure 15: Top 10 of Most Important Features in Predicting Divorce

- **Correlation with Divorce**:
    - Top Positively Correlated Questions

Code :

```
correlations =
df.corr()['Divorce'].drop('Divorce').sort_values(ascending=False)
print("Top 10 Positively Correlated Questions:\n",
correlations.head(10))
print("Top 10 Negatively Correlated Questions:\n",
correlations.tail(10))
```

Table 10: Top Positively Correlated Questions with Divorce

Top Positively Correlated Questions	Correlation
Q40	0.938
Q17	0.923
Q18	0.923
Q19	0.918
Q12	0.913
Q11	0.912
Q15	0.908
Q20	0.901
Q26	0.900
Q41	0.894

- **Top Negatively Correlated Questions**

Table 11: Top Negatively Correlated Questions with Divorce

Top Negatively Correlated Questions	Correlation
Q53	0.711 (less strongly positive)
Q29	0.669
Q47	0.654
Q52	0.651
Q28	0.566
Q46	0.444
Q6	0.430
Q7	0.429

- **Mean Score Comparison**:
    - Divorced performed better on average in these top-ranked attributes. This pattern will be investigated further during the conversation.

Code :

```
all_questions = [f'Q{i}' for i in range(1, 55)]
mean_scores = df.groupby('Divorce')[all_questions].mean().T
mean_scores.columns = ['Not Divorced (0)', 'Divorced (1)']
display(mean_scores.sort_values(by='Divorced (1)', ascending=False))
```

Table 12: Mean Score Comparison Between Divorced and Not Divorced Groups

Question	Not Divorced	Divorced
Q40	0.209	3.571
Q41	0.477	3.548
Q39	0.570	3.643
Q49	1.279	3.512
Q52	1.570	3.488

Code :

```
top_diff = mean_scores.sort_values(by='Divorced (1)',
ascending=False).head(5)
top_diff.plot(kind='barh')
plt.title("Top 5 Questions with Highest Mean Difference")
plt.xlabel("Average Response")
plt.show()
```

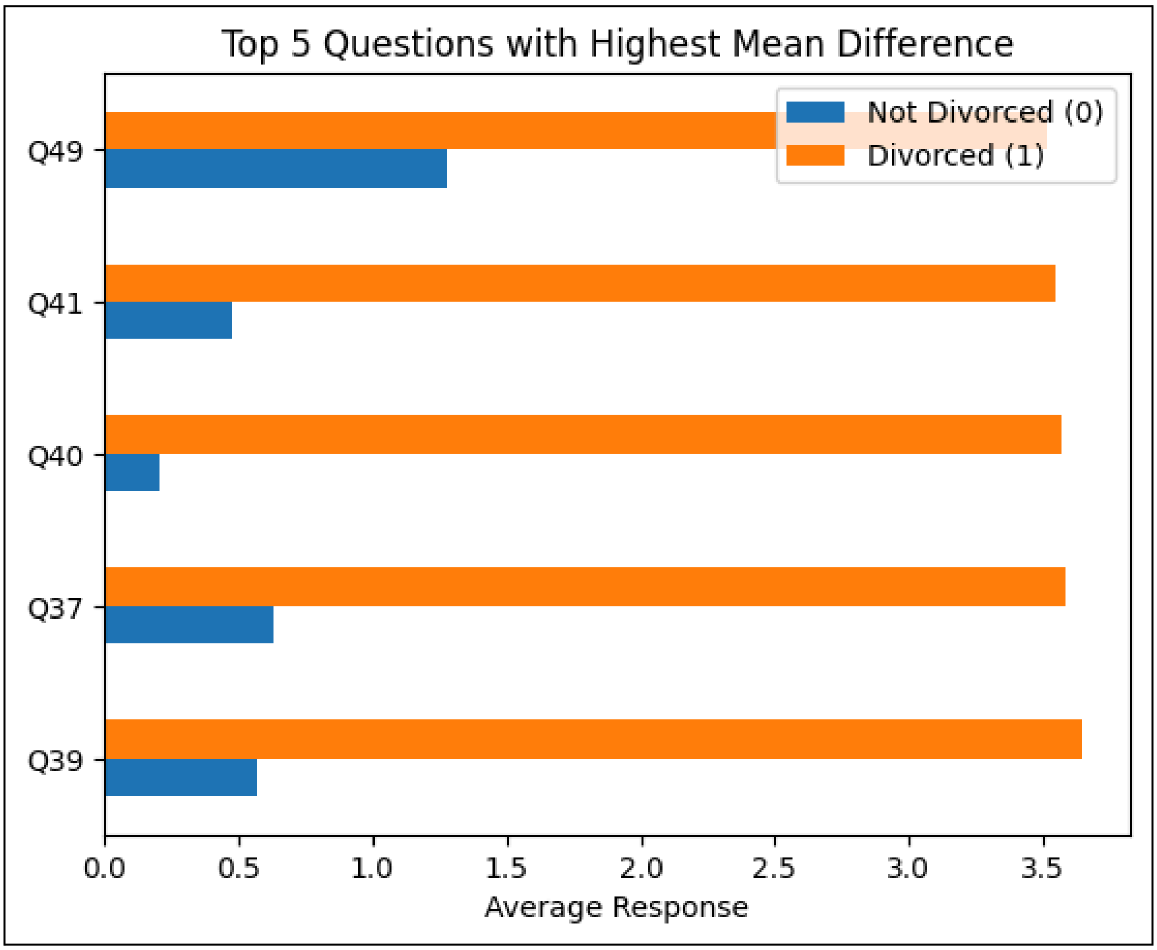

Figure 16: Top 5 Questions with Highest Mean Difference

DISCUSSIONS

The findings from both primary and secondary datasets provide insight into the complex relationship of factors impacting Muslim marriage success and divorce risk. This section critically examines the findings, compares the datasets, discusses limitations, and addresses the broader impact in the context of Malaysian Muslim society.

Primary Dataset

The primary dataset, compiled from a survey of 50 married Muslim adults in Malaysia, discovered several important indicators of marital stability. The Random Forest Classifier performed perfectly (Accuracy = 1.00, AUC = 1.00). Although this suggests significant patterns in the dataset, it also increases the likelihood of overfitting due to the small sample size and the removal of "Maybe" class entries to achieve binary classification.

ReligionCentrality (correlation = 0.91) and MarriageSatisfaction (0.90) emerged as the most strongly connected factors with perceived marriage success. This is consistent with the Islamic background in Malaysia, where religion plays a central role in forming family values, everyday life, and relationship expectations. Couples who shared strong Islamic beliefs and saw religion as important to their relationship reported higher levels of satisfaction and stability.

Some other important indicators are included as stable income, financial planning, communication frequency, and emotional well-being. These findings show that economic security and regular emotional support are essential components in maintaining long-term marriages. Financial preparation and the absence of frequent major arguments were also important in demonstrating the interconnection of practical, emotional, and spiritual components in marriage satisfaction.

Secondary Dataset

The secondary dataset is from Kaggle, which provides a more generalized psychological profile, concentrating on 170 people's responses to the Divorce Predictors Scale (DPS). The Random Forest model worked well (Accuracy = 0.97, AUC = 1.00), indicating that the chosen features were highly predictive of divorce outcomes.

Top predictive factors (e.g., Q40: Arguments, Q17: Emotional Disconnect, Q12: Mutual Understanding) had a major positive connection with divorce (correlations > 0.91). Mean score comparisons revealed that divorced people reported significantly higher levels of conflict, unhappiness, and lack of support. These signs support basic psychological facts about marriage, like emotional distance and unresolved conflict, which are key factors in marital breakdowns.

Comparison and Critical Reflection

Both datasets demonstrated the important role of communication quality, emotional connection, and financial stability in predicting marital outcomes. However, the primary dataset provided unique insights into the religious and cultural components that are especially important in the Malaysian Muslim context area which is frequently ignored in worldwide divorce research.

The secondary dataset showed strong psychological predictors, but it lacked cultural diversity, such as religious identity, family-related influence, and societal expectations. The difference emphasizes the importance of regional, culturally sensitive data collection while attempting to comprehend marriage in varied societies.

Despite their high accuracy scores, both models require critical reflection. The main dataset's perfect score could be attributed to its restricted size and class imbalance after filtering out "Maybe" responses. Furthermore, because the survey is self-reported, there is a risk of bias due to social desirability or individual misjudgment. Additionally, while the secondary dataset's questions are based on proven psychological scales, they may not completely capture the religious and cultural dynamics unique to Muslim spouses in Southeast Asia.

Implications and Practical Use

These findings have serious consequences. Counselors and Islamic family institutions can use the discovered variables (such as Religion Centrality, Finance Planning, and Communication Frequency) to better personalize premarital and marital counseling programs. Policymakers may implement targeted interventions, such as financial planning workshops and religious compatibility modules, into mandated premarital courses (e.g., Kursus Kahwin). Data-driven solutions might also be created for early risk screening, allowing couples and advisors to address significant areas of concern before disagreements escalate.

FUTURE WORKS

While this study successfully identified significant factors influencing Muslim marriage success using machine learning and statistical analysis, there are several opportunities to enhance and expand the research in the future:

Increase Sample Size for Primary Data

The current primary dataset consisted of only 50 valid responses, which, while insightful, limits the generalizability of the findings. Future studies should aim to collect data from a larger and more diverse group of Muslim respondents across different regions in Malaysia to improve model robustness and representativeness.

Include Longitudinal Data

The current study used cross-sectional survey data, capturing only a snapshot of marital conditions. Future work can incorporate longitudinal tracking to observe changes in marital satisfaction over time, which would allow for time-series modeling and more accurate prediction of long-term outcomes.

Cross-Cultural Comparative Studies

Conducting similar studies among Muslim populations in other countries (Indonesia, Pakistan, or the Middle East) would allow for cross-cultural comparison and highlight whether predictors of marital success vary across different Islamic societies.

Build a Publicly Accessible Dashboard

As an applied outcome, the findings could be integrated into a user-friendly web dashboard or mobile application that helps individuals and counselors assess relationship health using the trained machine learning model.

CONCLUSION

This study managed to identify the key success and failure factors of Muslim marriages through the comparison of both primary and secondary datasets by using data science techniques. The findings of the primary dataset, which were gathered from Malaysian Muslim couples, revealed that religiosity, financial security, mental well-being, and proper communication are essential for marriage stability. Plots such as correlation heatmaps and feature importance plots readily identified ReligionCentrality, StableIncome, and MarriageSatisfaction as the most predictive features of a successful Malaysian Muslim marriage.

The secondary data set, drawn from responses to the Divorce Predictors Scale answers, supported these findings from a more general psychological perspective. It showed that regular conflict, emotional distance, and lack of mutual understanding were the most significant divorce predictors. Despite cultural differences, both data sets emphasized the importance of communication and emotional support for good relationships.

The Random Forest Classifier, one of the machine learning models, was extremely precise in predicting marriage outcomes in both datasets. The flawless performance with the primary dataset points toward possible overfitting due to the small sample size. However, the study reveals the success of combining statistical analysis and machine learning in uncovering actionable insights.

Lastly, this research provides policymakers, religious leaders, and counselors with an evidence-based foundation to build stronger, more culturally appropriate marriage-strengthening initiatives. With emotional, financial, and religious aspects combined, early interventions can better target marital issues, reduce divorce rates, and promote healthier Muslim families in Malaysia and beyond.

REFERENCES

Pal, M., & Smita Parija. (2021). Prediction of Heart Diseases using Random Forest. *Journal of Physics Conference Series*, *1817*(1), 012009–012009. https://doi.org/10.1088/1742-6596/1817/1/012009

Sharma, A., Arshdeep Singh Chudhey, & Singh, M. (2021). *Divorce case prediction using Machine learning algorithms*. 214–219. https://doi.org/10.1109/icais50930.2021.9395860

Mian, Raza, A., Zhao, N., Tariq, A., Younas, F., Ahmed, G., Ullah, S., Syeda Fizzah Jillani, Abbas, I., & Aslam, M. (2022). Predicting Divorce Prospect Using Ensemble Learning: Support Vector Machine, Linear Model, and Neural Network. *Computational Intelligence and Neuroscience*, *2022*, 1–15. https://doi.org/10.1155/2022/3687598

Abdelkader Moumen, Shafqat, A., Tariq Alraqad, Etaf Saleh Alshawarbeh, Saber, H., & Ramsha Shafqat. (2024). Divorce prediction using machine learning algorithms in Ha'il region, KSA. *Scientific Reports*, *14*(1). https://doi.org/10.1038/s41598-023-50839-1

Aimran, N., Rambli, A., Afthanorhan, A., Mahmud, A., Sapri, A., & Aireen, A. (2022). Prediction of Malaysian women divorce using machine learning techniques / Nazim Aimran … [et al.] - UiTM Institutional Repository. *Uitm.edu.my*. https://ir.uitm.edu.my/id/eprint/69241/1/69241.pdf

Milani, A. S., Hosseini, M., Matbouei, F., & Nasiri, M. (2020). Effectiveness of emotional intelligence training program on marital satisfaction, sexual quality of life, and psychological well-being of women. *Journal of Education and Health Promotion, 9*, 149. https://doi.org/10.4103/jehp.jehp_124_20

Biggiogera, J., Paleari, F. G., Regalia, C., & Fincham, F. D. (2021). Predicting communication behavior during couples' conflict using BERT and LIWC features. *Journal of Social and Personal Relationships, 38*(11), 3103–3126. https://cocoa.ethz.ch/downloads/2022/01/2623_Biggiogera%20et%20al%202021%20BERT%20LIWC%20Couple%20Conflict.pdf

Vowels, L. M., Mark, K. P., & Leonard, L. M. (2021). What predicts sexual desire? A machine learning approach to understanding desire in romantic relationships. *The Journal of Sexual Medicine, 18*(8), 1364–1376. https://doi.org/10.1016/j.jsxm.2021.04.010

Azhar, N. I., & Mohd Hoesni, S. (2023). Hubungan antara kecerdasan emosi dengan kepuasan perkahwinan sewaktu pandemik COVID‑19 di Malaysia. *e-Bangi: Journal of Social Sciences and Humanities, 20*(1), 406–417. http://dx.doi.org/10.17576/ebangi.2023.2001.32

Chen, Q., Zhang, M., Wu, W., Liu, D., Liu, T., & Yao, Y. (2024). The effect of marital status on survival in pancreatic ductal adenocarcinoma: A population-based study. *Scientific Reports, 14*, Article 53145. https://doi.org/10.1038/s41598-024-53145-6