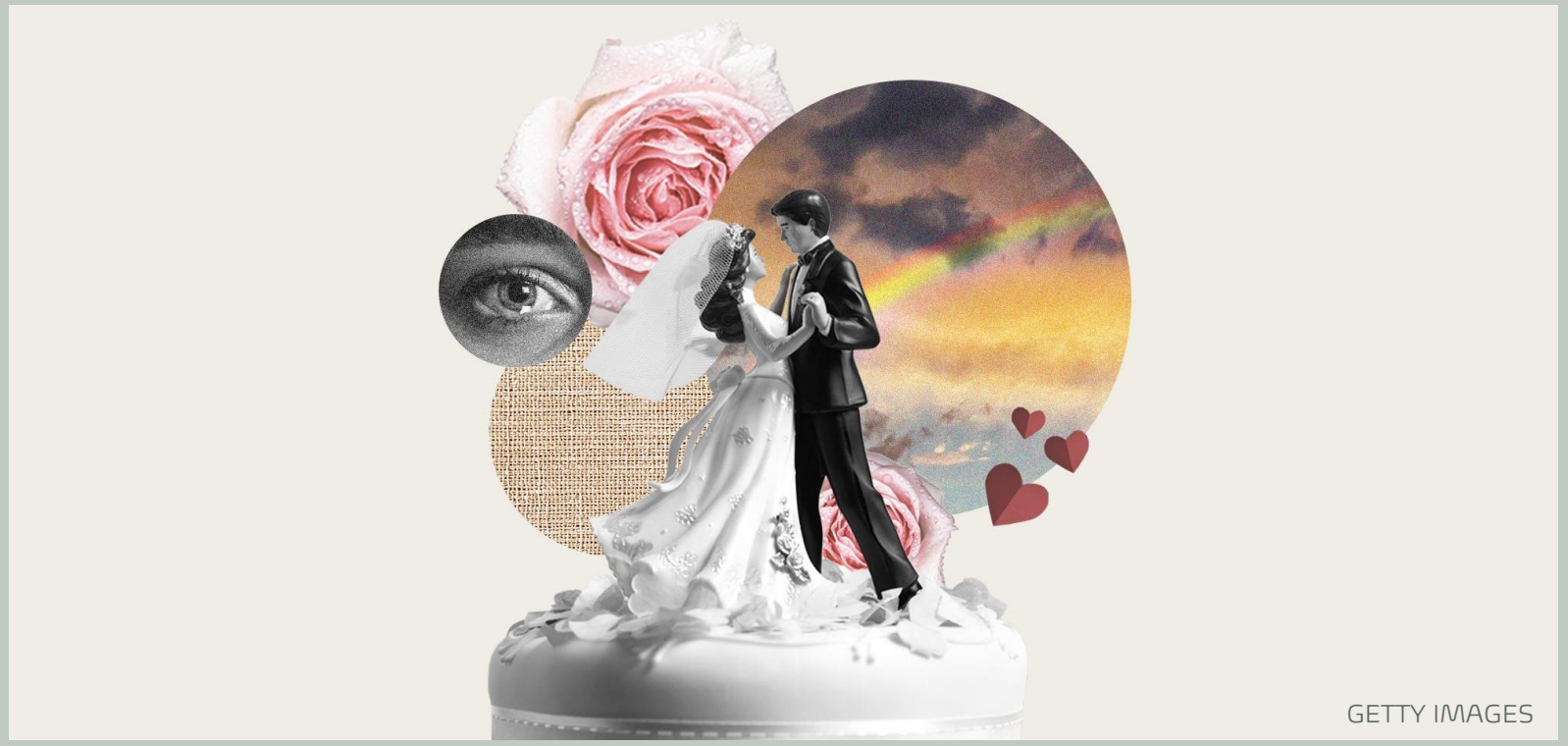
Divorce Predictors



GETTY IMAGES

Prepared by: Cheryl Lim

Agenda

01 Introduction

02 Methodology

03 Process Workflow

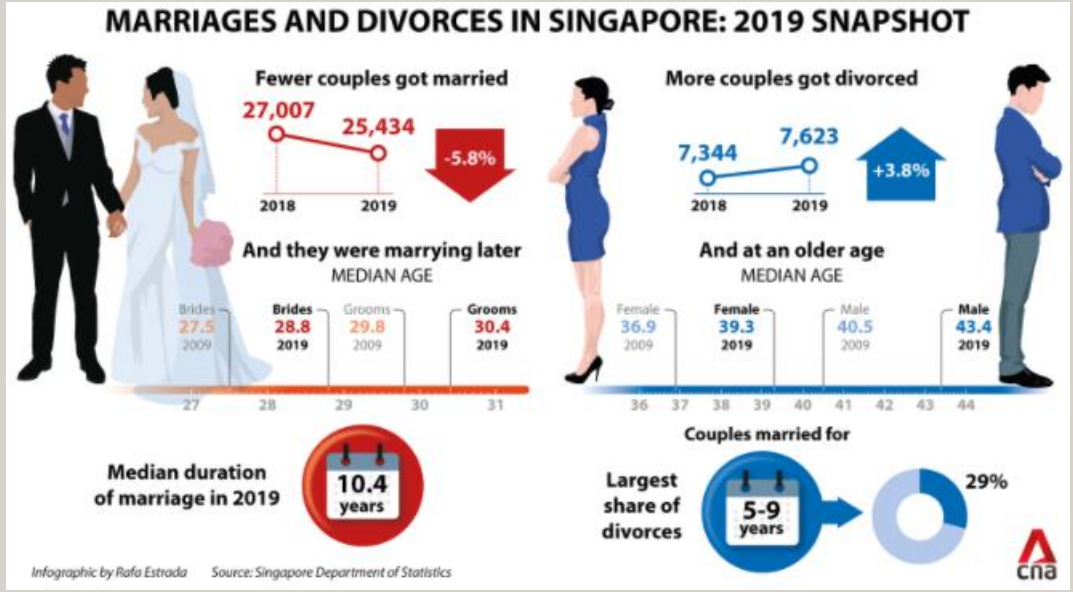
04 Results

05 Conclusions

06 Future Opportunities

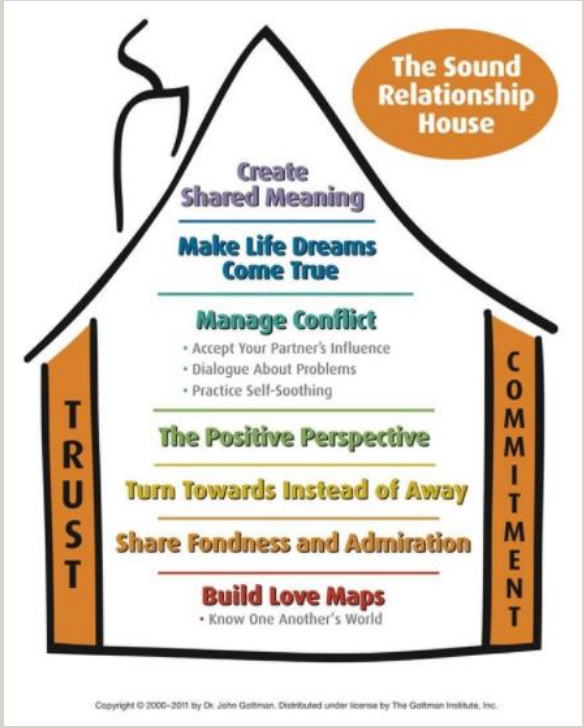
01 Introduction

- For this project, I am an employee in the Ministry of Social and Family Development
- Presenting to internal management and external organizations/ professionals providing marriage counselling services
- There is a need to identify the important factors leading to divorce, which could influence the formulation of policies and counselling practices
- Many divorces can be prevented if predicted early



01 Introduction

- A Divorce Predictor Scale (DPS) was developed on the basis of the Gottman Method Couples Therapy
- The aim of this project is to determine the predictive power of DPS and identify the important factors leading to divorce



Images from the Gottman Institute

02 Methodology

Divorce Predictors Data Set

UCI Machine Learning Repository

https://archive.ics.uci.edu/ml/datasets/Divorce+Predictors+data+set#

Data Set Characteristics:	Multivariate, Univariate	**Number of Instances:**	170	**Area:**	Life
Attribute Characteristics:	Integer	**Number of Attributes:**	54	**Date Donated**	2019-07-24
Associated Tasks:	Classification	**Missing Values?**	N/A	**Number of Web Hits:**	99298

Divorce Predictor Scale (DPS)

(answered on a scale of 0 – strongly disagree to 4 – strongly agree)

1. If one of us apologizes when our discussion deteriorates, the discussion ends.
2. I know we can ignore our differences, even if things get hard sometimes.
3. When we need it, we can take our discussions with my spouse from the beginning and correct it.
4. When I discuss with my spouse, to contact him will eventually work.
5. The time I spent with my wife is special for us.
6. We don't have time at home as partners.
7. We are like two strangers who share the same environment at home rather than family.
8. I enjoy our holidays with my wife.
9. I enjoy traveling with my wife.
10. Most of our goals are common to my spouse.
11. I think that one day in the future, when I look back, I see that my spouse and I have been in harmony with each other.
12. My spouse and I have similar values in terms of personal freedom.
13. My spouse and I have similar sense of entertainment.
14. Most of our goals for people (children, friends, etc.) are the same.
15. Our dreams with my spouse are similar and harmonious.
16. We're compatible with my spouse about what love should be.
17. We share the same views about being happy in our life with my spouse
18. My spouse and I have similar ideas about how marriage should be
19. My spouse and I have similar ideas about how roles should be in marriage
20. My spouse and I have similar values in trust.
21. I know exactly what my wife likes.
22. I know how my spouse wants to be taken care of when she/he sick.
23. I know my spouse's favorite food.
24. I can tell you what kind of stress my spouse is facing in her/his life.
25. I have knowledge of my spouse's inner world.
26. I know my spouse's basic anxieties.
27. I know what my spouse's current sources of stress are.
28. I know my spouse's hopes and wishes.
29. I know my spouse very well.
30. I know my spouse's friends and their social relationships.
31. I feel aggressive when I argue with my spouse.
32. When discussing with my spouse, I usually use expressions such as 'you always' or 'you never' .
33. I can use negative statements about my spouse's personality during our discussions.
34. I can use offensive expressions during our discussions.
35. I can insult my spouse during our discussions.
36. I can be humiliating when we discussions.
37. My discussion with my spouse is not calm.
38. I hate my spouse's way of open a subject.
39. Our discussions often occur suddenly.
40. We're just starting a discussion before I know what's going on.
41. When I talk to my spouse about something, my calm suddenly breaks.
42. When I argue with my spouse, ı only go out and I don't say a word.
43. I mostly stay silent to calm the environment a little bit.
44. Sometimes I think it's good for me to leave home for a while.
45. I'd rather stay silent than discuss with my spouse.
46. Even if I'm right in the discussion, I stay silent to hurt my spouse.
47. When I discuss with my spouse, I stay silent because I am afraid of not being able to control my anger.
48. I feel right in our discussions.
49. I have nothing to do with what I've been accused of.
50. I'm not actually the one who's guilty about what I'm accused of.
51. I'm not the one who's wrong about problems at home.
52. I wouldn't hesitate to tell my spouse about her/his inadequacy.
53. When I discuss, I remind my spouse of her/his inadequacy.
54. I'm not afraid to tell my spouse about her/his incompetence.

170 Turkish Couples

Class 0 (49%) – Divorced

Class 1 (51%) – Married, without any thoughts of divorce

02 Methodology

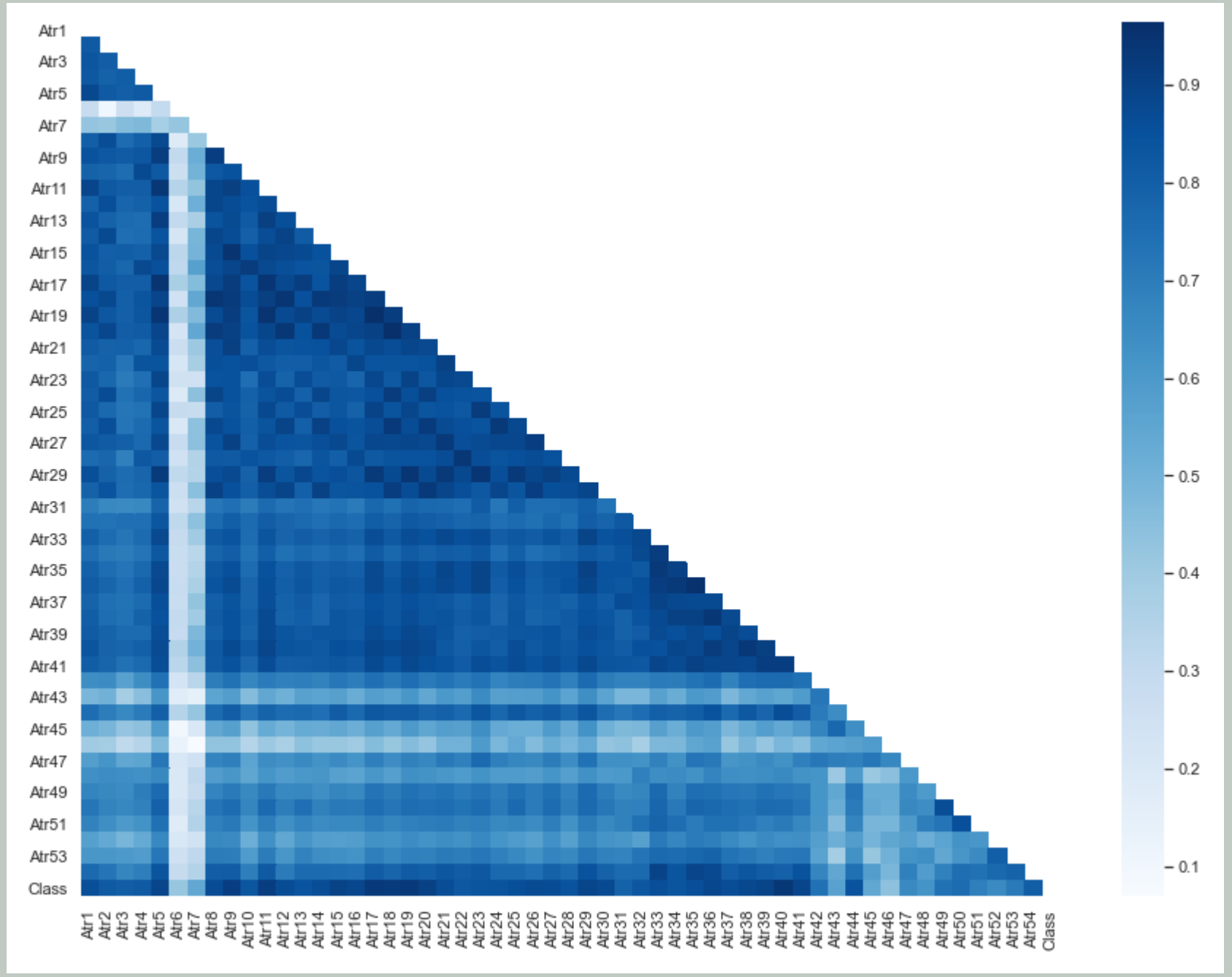
Models:	(1) Logistic Regression
	(2) Random Forest
	(3) Multilayer Perceptron
Metrics:	F1 Score/ Accuracy
Tools:	Python

03 Process Workflow

EDA (Correlation Matrix)

TOP 6 Features of Highest Correlation with Class:

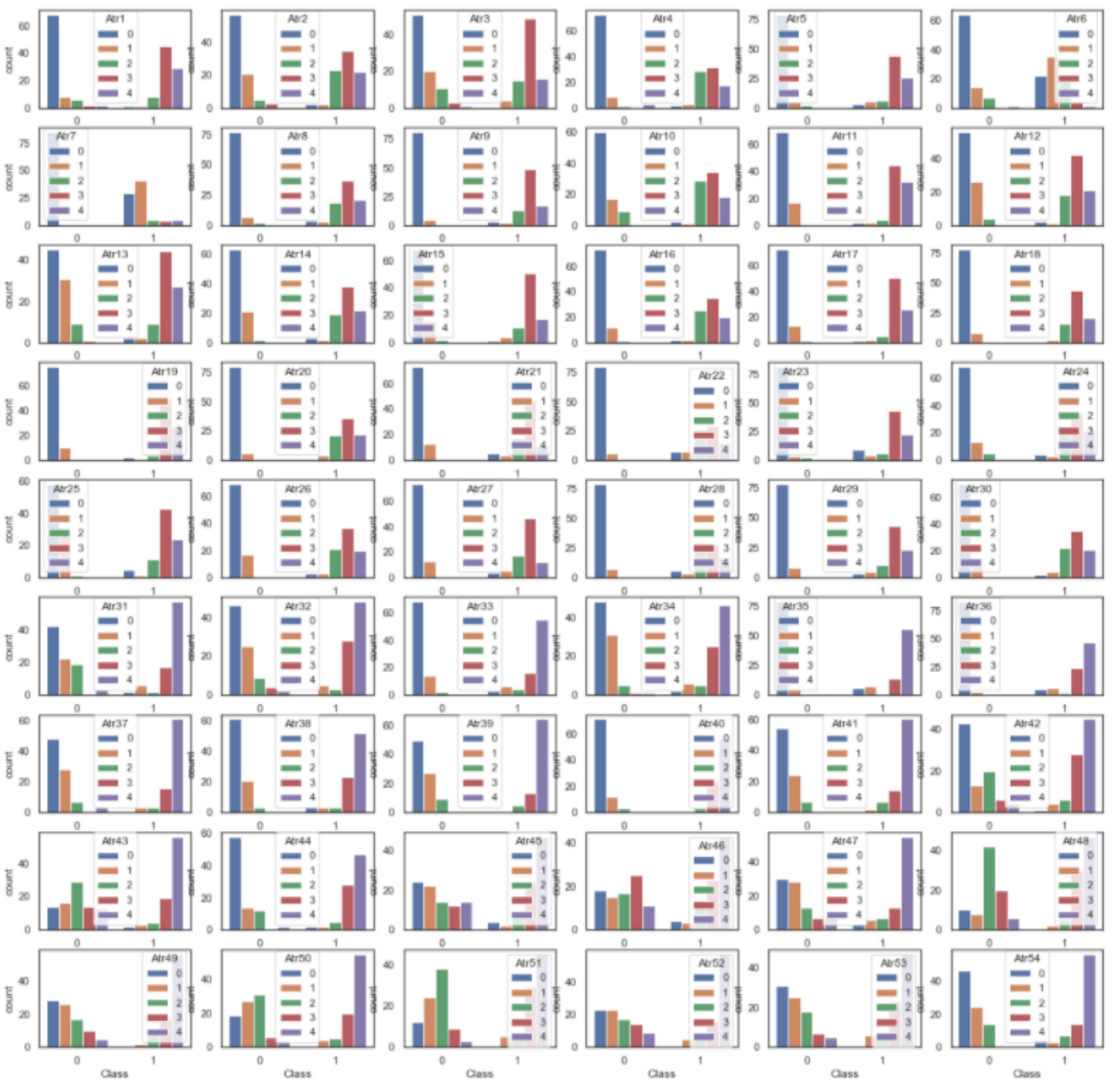
Feature	Correlation Coefficient
Atr40	0.9387
Atr17	0.9293
Atr19	0.9286
Atr18	0.9232
Atr11	0.9184
Atr9	0.9124



EDA (Count Plot)

Positive Statements:
Married tend to agree

Negative Statements:
Married tend to disagree



Hyperparameter Tuning

```
# Logistic Regression
logreg = LogisticRegression(n_jobs=-1)
```

```
%%time
# Hyperparameter tuning using K-fold cross validation
# ... via Grid Search method
param_grid = {'penalty': ['l1', 'l2', 'elasticnet', 'none'],
              'C': np.logspace(-2, 2, 5)}

gs_logreg = GridSearchCV(logreg,
                         param_grid,
                         cv=5,
                         scoring='f1_macro',
                         n_jobs=-1)
gs_logreg.fit(X_train_scaled, y_train)
```

```
Wall time: 4.5 s

GridSearchCV(cv=5, estimator=LogisticRegression(n_jobs=-1), n_jobs=-1,
             param_grid={'C': array([1.e-02, 1.e-01, 1.e+00, 1.e+01, 1.e+02]),
                         'penalty': ['l1', 'l2', 'elasticnet', 'none']},
             scoring='f1_macro')
```

```
# Best model hyperparameters and score
print(gs_logreg.best_estimator_)
print(gs_logreg.best_params_)
print(gs_logreg.best_score_)
```

```
LogisticRegression(C=0.01, n_jobs=-1)
{'C': 0.01, 'penalty': 'l2'}
0.9839484702093397
```

	Model	Training Acc	Testing Acc	Precision	Recall	F1 Score
0	Logistic Regression	0.984252	0.953488	0.958333	0.952381	0.953261
1	Random Forest	1.000000	0.953488	0.958333	0.952381	0.953261
2	Multilevel Perceptron	0.984252	0.953488	0.958333	0.952381	0.953261

All 3 models give the same scores

```
# Random Forest
rf = RandomForestClassifier(random_state=0)
```

```
%%time
# Hyperparameter tuning using K-fold cross validation
# ... via Grid Search method
param_grid = {'n_estimators': [100, 200, 300],
              'criterion': ['gini', 'entropy'],
              'max_depth': [5, 10, 25],
              'min_samples_split': [5, 10, 15]}

gs_rf = GridSearchCV(rf,
                     param_grid,
                     cv=5,
                     scoring='f1_macro',
                     n_jobs=-1)
gs_rf.fit(X_train_scaled, y_train)
```

```
Wall time: 42.9 s

GridSearchCV(cv=5, estimator=RandomForestClassifier(random_state=0), n_jobs=-1,
             param_grid={'criterion': ['gini', 'entropy'],
                         'max_depth': [5, 10, 25],
                         'min_samples_split': [5, 10, 15],
                         'n_estimators': [100, 200, 300]},
             scoring='f1_macro')
```

```
# Best model hyperparameters and score
print(gs_rf.best_estimator_)
print(gs_rf.best_params_)
print(gs_rf.best_score_)
```

```
RandomForestClassifier(max_depth=5, min_samples_split=5, random_state=0)
{'criterion': 'gini', 'max_depth': 5, 'min_samples_split': 5, 'n_estimators': 100}
0.9839484702093397
```

```
# Multi-Level Perceptron (stochastic iterative)
mlp = MLPClassifier(solver='sgd')
```

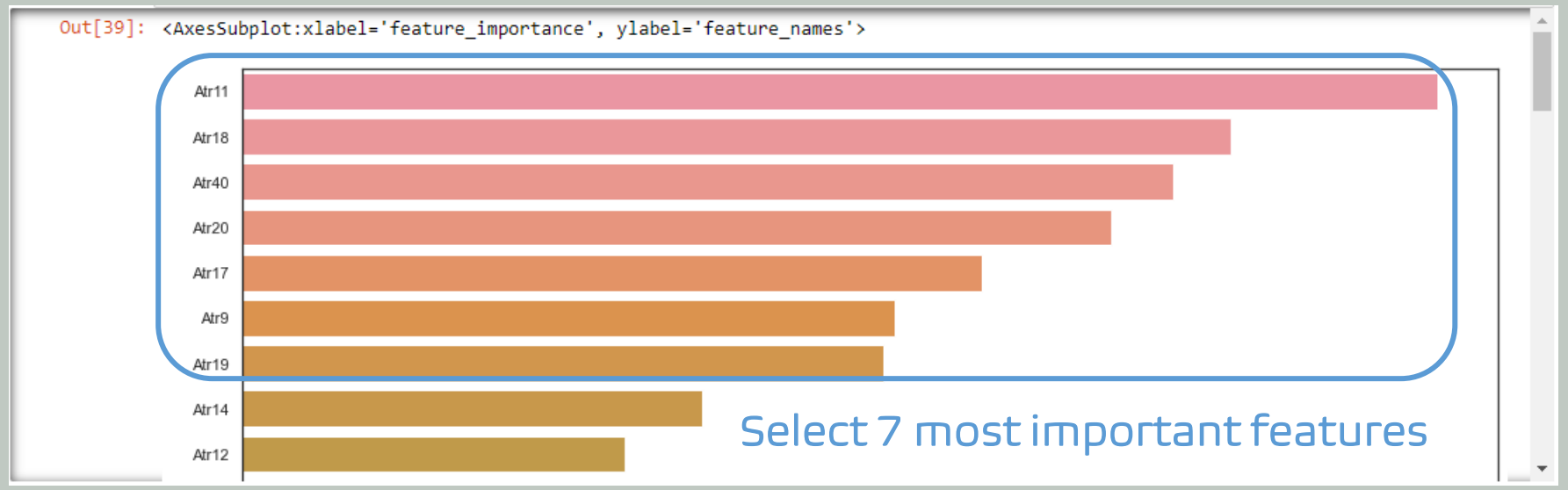
```
%%time
# Hyperparameter tuning using K-fold cross validation
# ... via Grid Search method
param_grid = {'hidden_layer_sizes': [(3,3),
                                     (2),
                                     (3)],
              'alpha': np.logspace(-4, -1, 4),
              'max_iter': [400, 450, 500, 550]}

gs_mlp = GridSearchCV(mlp,
                      param_grid,
                      cv=5,
                      scoring='f1_macro',
                      n_jobs=-1)
gs_mlp.fit(X_train_scaled, y_train)
```

```
Wall time: 25.5 s

C:\Users\user\anaconda3\lib\site-packages\sklearn\neural_network\_multilayer_perceptron.py:582: ConvergenceWarning: Stochastic
Optimizer: Maximum iterations (550) reached and the optimization hasn't converged yet.
  warnings.warn(

GridSearchCV(cv=5, estimator=MLPClassifier(solver='sgd'), n_jobs=-1,
             param_grid={'alpha': array([0.0001, 0.001 , 0.01  , 0.1   ]),
                         'hidden_layer_sizes': [(3, 3), 2, 3],
                         'max_iter': [400, 450, 500, 550]},
             scoring='f1_macro')
```

```
# Best model hyperparameters and score
print(gs_mlp.best_estimator_)
print(gs_mlp.best_params_)
print(gs_mlp.best_score_)
```

```
MLPClassifier(hidden_layer_sizes=2, max_iter=550, solver='sgd')
{'alpha': 0.0001, 'hidden_layer_sizes': 2, 'max_iter': 550}
0.9919484702093397
```

Feature Selection (Method 1)



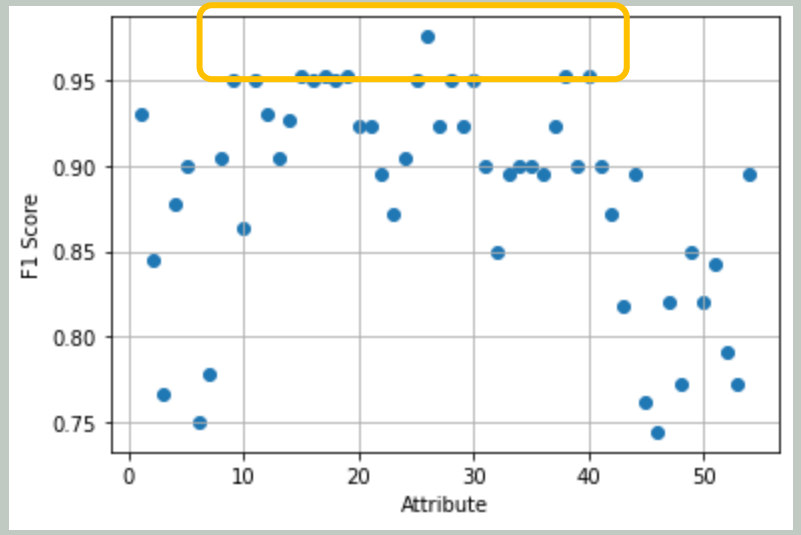
	Model	Training Acc	Testing Acc	Precision	Recall	F1 Score
0	Logistic Regression	1.000000	0.953488	0.958333	0.952381	0.953261
1	Random Forest	1.000000	0.953488	0.958333	0.952381	0.953261
2	Multilevel Perceptron	0.992126	0.930233	0.931522	0.929654	0.930081

No increase in scores

Feature Selection (Method 2)

ML model training on one feature at a time:

- 1 feature (Atr26) gave the highest F1 score of 0.9756
- 5 features (Atr15, Atr17, Atr19, Atr38, Atr40) each gave the same F1 score of more than 0.9500



	Model	Training Acc	Testing Acc	Precision	Recall	F1 Score
0	Logistic Regression	1.000000	0.976744	0.978261	0.976190	0.976694
1	Random Forest	0.992126	0.976744	0.978261	0.976190	0.976694
2	Multilevel Perceptron	0.984252	0.953488	0.958333	0.952381	0.953261

Higher scores!

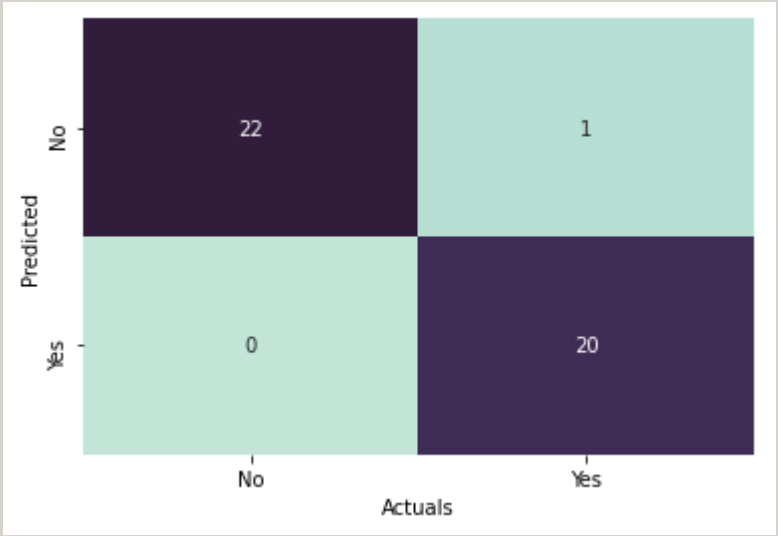
04 Results

- Logistic Regression was selected because it gave the highest score within the shortest time

Classification Report

	precision	recall	f1-score	support
0	0.96	1.00	0.98	22
1	1.00	0.95	0.98	21
accuracy			0.98	43
macro avg	0.98	0.98	0.98	43
weighted avg	0.98	0.98	0.98	43

Confusion Matrix



05 Conclusions

- The aim of this project was to determine the predictive power of DPS and identify the important factors leading to divorce
- The machine learning model can predict divorce with a high accuracy of 97.67%
- The important factors leading to divorce have been identified as follows:

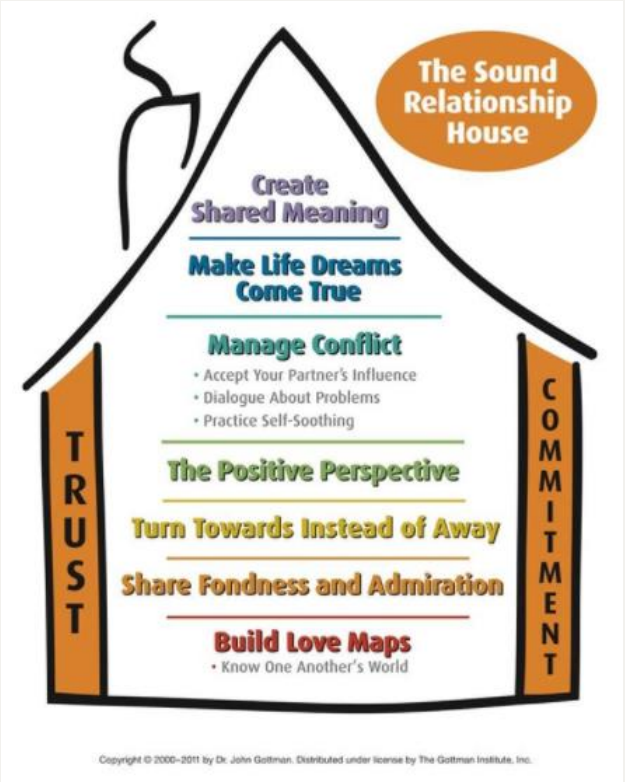


Image from the Gottman Institute

Create Shared Meaning:

15. Our dreams with my spouse are similar and harmonious.

17. We share the same views about being happy in our life with my spouse

19. My spouse and I have similar ideas about how roles should be in marriage

Manage Conflict:

38. I hate my spouse's way of open a subject.

40. We're just starting a discussion before I know what's going on.

Build Love Maps:

26. I know my spouse's basic anxieties.

GETTY IMAGES

Q&A