



Love Connection

QTM 347 Final Project

Authors: Natasha Gandhi & Ishana Ram

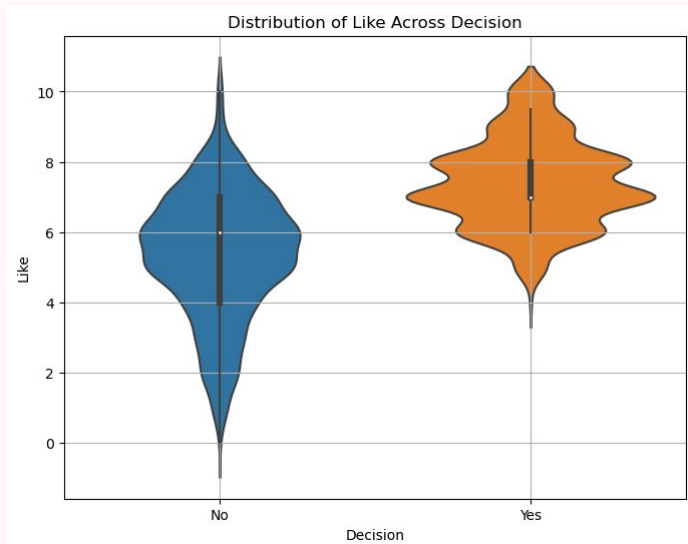
Motivation

In this project, we're exploring the dynamics of human compatibility and attraction in relationships. We are using speed-dating data to better understand the qualities and personalities that foster connections and mutual attraction. As young adults navigating modern relationships, we hope to gain insights that can enrich our own experiences and help us build more meaningful and fulfilling relationships.

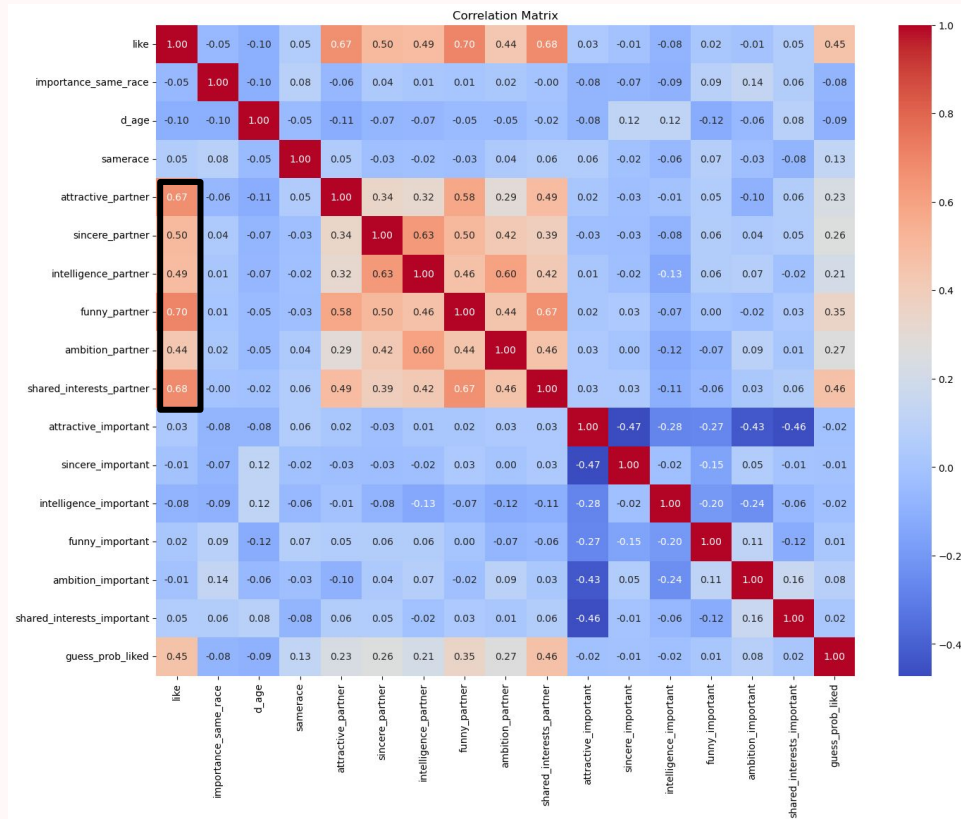
We intend to conduct this analysis by using the following models:

- Random Forest with Bagging
- Best Subset Selection
- Lasso Regularization
- Decision Tree Regressor

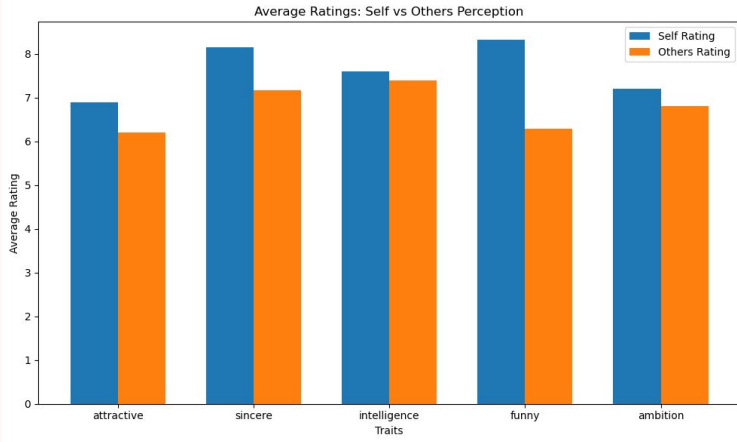
Introductory Data Analysis



People's decisions are sometimes unpredictable!

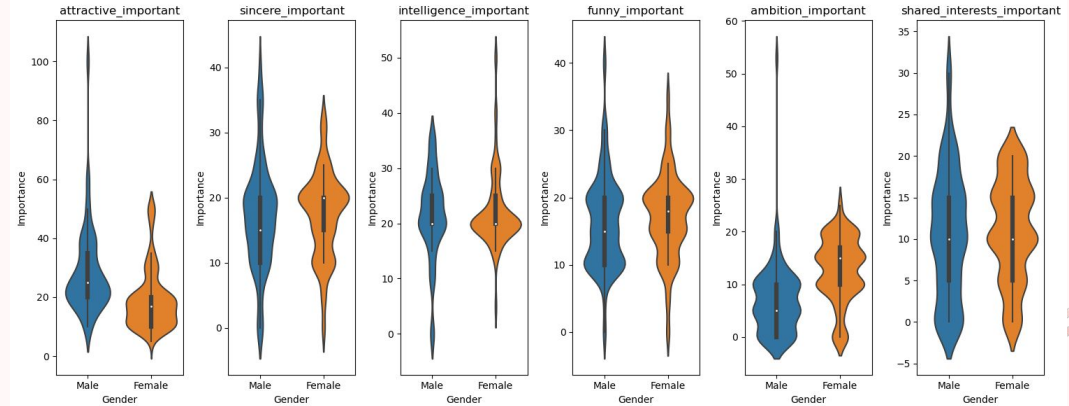


Interesting Findings

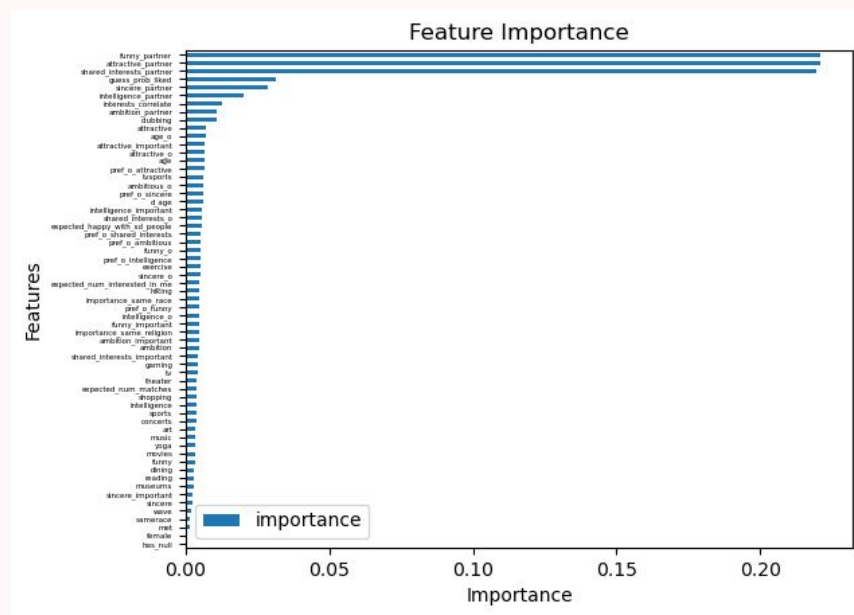
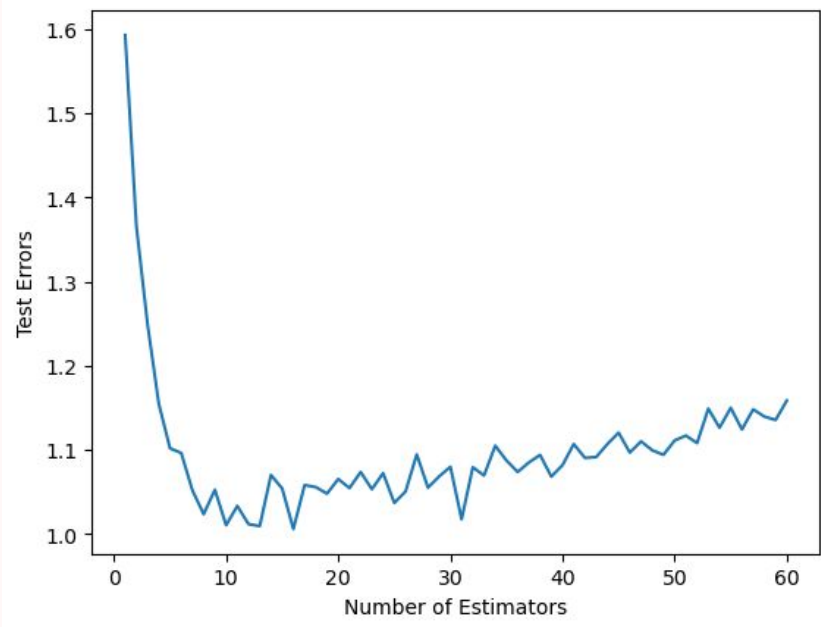


People tend to rate themselves higher than their partner rates them across characteristic ratings!

On average, men tend to place higher importance on partner attractiveness, and women tend to place higher importance on partner sincerity, humor, and ambition



Random Forest with Bagging



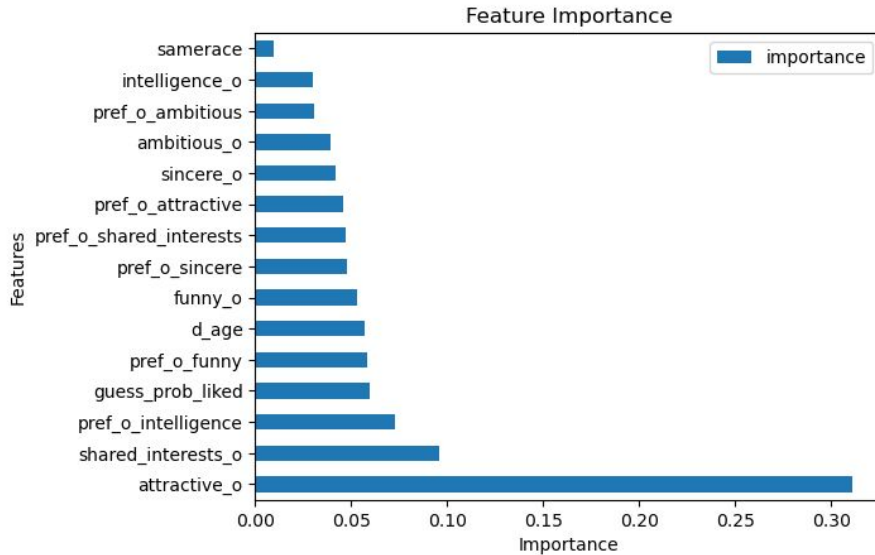
Lowest Test MSE ($m = 16$) = 1.0058558333333334

	importance
funny_partner	0.179467
attractive_partner	0.178709
shared_interests_partner	0.155488
sincere_partner	0.057959
guess_prob_liked	0.051337
intelligence_partner	0.035181
ambition_partner	0.023904
interests_correlate	0.012525
expected_num_matches	0.012295
shared_interests_o	0.009193
met	0.008970
clubbing	0.008428
attractive	0.008120
pref_o_attractive	0.008111
ambition	0.008036
age_o	0.007992
tvsports	0.007752
attractive_o	0.007666
age	0.007250
funny_o	0.007187
expected_happy_with_sd_people	0.006910
pref_o_shared_interests	0.006869
ambitious_o	0.006788
d_age	0.006726
intelligence_o	0.006698
pref_o_sincere	0.006691
expected_num_interested_in_me	0.006642
pref_o_ambitious	0.006621
attractive_important	0.006550
importance_same_race	0.006469
shopping	0.006227
sincere_o	0.006107
pref_o_intelligence	0.006077

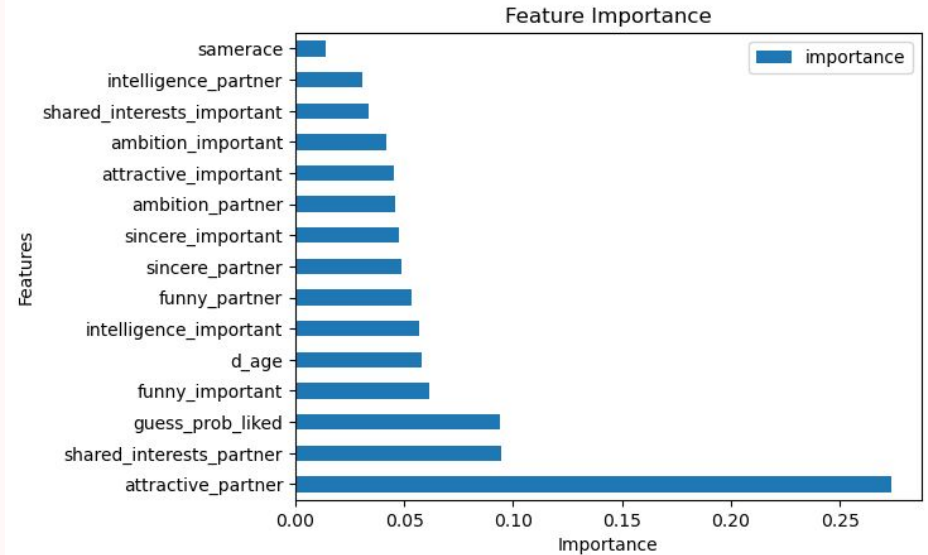
Top Features

- Top features include funny_partner, attractive_partner, and shared_interests_partner
- Features such as tv, hiking, art, shopping and other interest variables were at the bottom

Feature Importance: Decision vs. Decision_o



(M = 7): Training MSE: 0.01863305489260143
Test MSE: 0.13936809523809524



(M = 9): Training MSE: 0.0186909307875895
Test MSE: 0.11304761904761904

Best Subset Selection

- We used best subset selection to determine the top three predictors of 'like'
- We randomly split the data into 80% training and 20% testing data

Results:

- Training MSE: 1.1534185844579106
- Test MSE: 1.368364626343596
- Top three predictors of like:
 - 1. attractive_partner
 - 2. shared_interests_partner
 - 3. funny_partner

OLS Regression Results

Dep. Variable:	like	R-squared:	0.669			
Model:	OLS	Adj. R-squared:	0.668			
Method:	Least Squares	F-statistic:	562.2			
Date:	Sun, 28 Apr 2024	Prob (F-statistic):	7.66e-200			
Time:	13:45:38	Log-Likelihood:	-1250.4			
No. Observations:	839	AIC:	2509.			
Df Residuals:	835	BIC:	2528.			
Df Model:	3					
Covariance Type:	nonrobust					
	coef	std err	t	P> t	[0.025	0.975]
attractive_partner	0.3509	0.025	14.210	0.000	0.302	0.399
funny_partner	0.2404	0.027	9.041	0.000	0.188	0.293
shared_interests_partner	0.3020	0.024	12.818	0.000	0.256	0.348
const	0.8664	0.140	6.189	0.000	0.592	1.141
Omnibus:	62.075	Durbin-Watson:	1.980			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	180.188			
Skew:	-0.345	Prob(JB):	7.46e-40			
Kurtosis:	5.163	Cond. No.	41.3			

Lasso Regularization

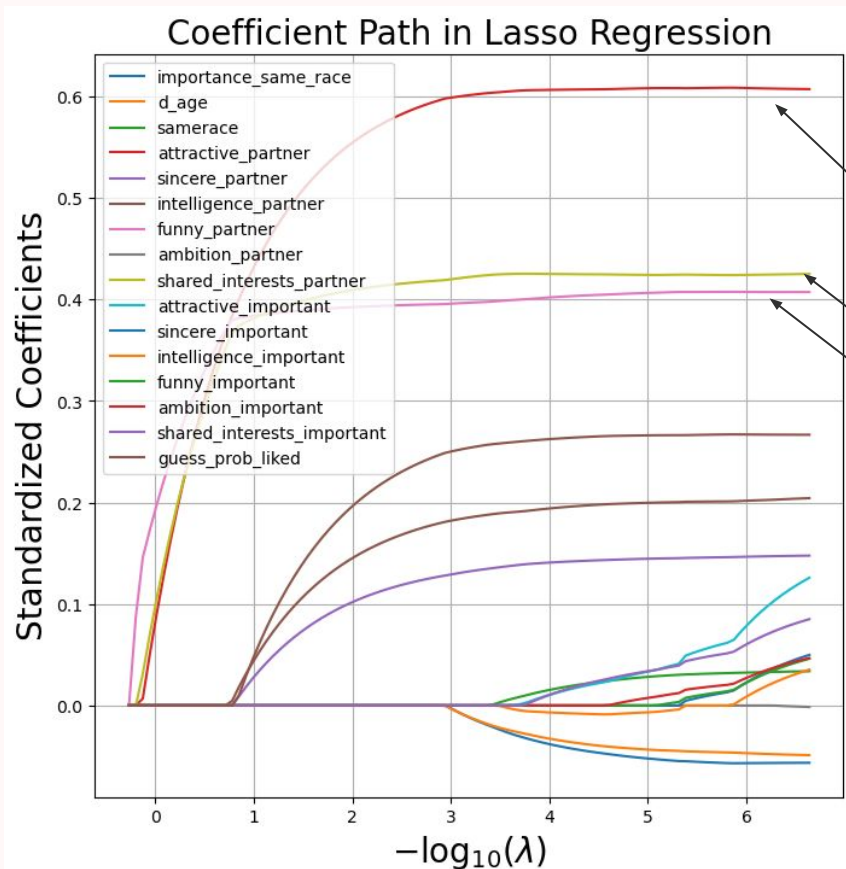
Feature	Coefficient
attractive_partner	0.622564
shared_interests_partner	0.480673
funny_partner	0.367187
guess_prob_liked	0.240499
intelligence_partner	0.190104
sincere_partner	0.135778
attractive_important	0.006206
d_age	-0.000000
samerace	0.000000
ambition_partner	0.000000
sincere_important	0.000000
funny_important	-0.000000
ambition_important	-0.000000
shared_interests_important	0.000000
intelligence_important	-0.010303
importance_same_race	-0.025238

- The data was standardized using `StandardScaler()` in the Pipeline
- CV was run using `skl.ElasticNetCV` to find the best alpha
- Lasso was fit to the training data

Results:

- Features with the most positive coefficients included `attractive_partner`, `shared_interests_partner`, and `funny_partner`
- Training MSE: 1.0312009921697962
- Test MSE: 1.1564421955496138

Coefficient Path in Lasso Regression

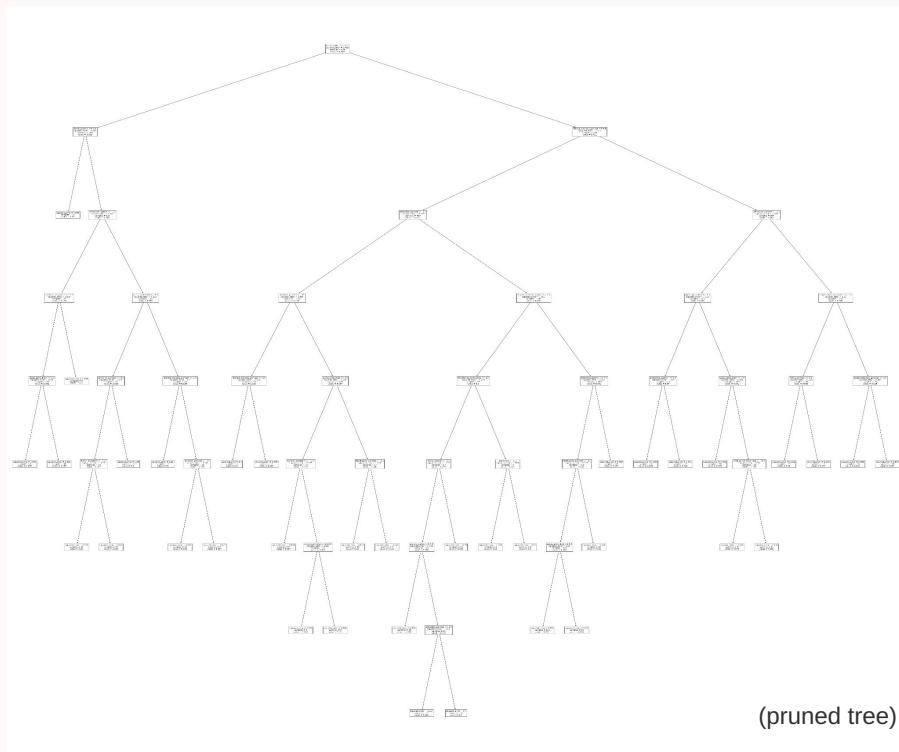


- Here we can see the path of the coefficients as alpha increases.
- The highest coefficients (top 3 lines) are once again attractive_partner, funny_partner, and shared_interests_partner

Decision Tree Regressor

The final step of our project was to create a decision tree based on the dataset. Ultimately, our goal was to transform this into an interactive algorithm that could be used by our peers.

1. First we fit a decision tree to our training set and evaluated the model on our testing set
 - a. Test MSE = 2.2117224880382773
2. We then pruned our tree in hopes of improving MSE
 - a. Test MSE = 1.710249225443976 ↓
 - b. Top Variables: funny_partner, attractive_partner, and shared_interest_partner



Model Comparison

Random Forest *

Training MSE: 0.0186

Test MSE: 0.1393

Lasso •

Training MSE: 1.0312

Test MSE: 1.1564

Best Subset Selection •

Training MSE: 1.1534

Test MSE: 1.3684

Pruned Decision Tree •

Training MSE: 0.7582

Test MSE: 1.7102

* used decision as the target variable

• used like as the target variable

Decision Tree Algorithm

Finally, we wanted to incorporate the results of our pruned decision tree into an interactive web application to provide information about how much someone may like their partner!

Using streamlit and a rendering platform we created an application that asks users can answer a series of questions about their personal preferences and opinions of their partner, similar to those asked in the speed dating event! When users click “calculate”, the algorithm will follow the decision tree and return a like score, which is the “value” on the decision tree.

Link: <https://qtm-347-project-2.onrender.com/>

Partner Liking Calculator

Please answer the following questions:

What is the age difference between you and your partner?

0.00

Are you and your partner of the same race?

☒ Yes

☐ No

How important is it to you that your partner is of the same race? (1-10)

1
1 10

On a scale of 1-10, how attractive do you find your partner?

1
1 10

On a scale of 1-10, how sincere do you find your partner?

1
1 10

On a scale of 1-10, how intelligent do you find your partner?

1
1 10



Thanks!

Any Questions?

Credits: This presentation template was created by [Slidesgo](#), and includes icons by [Flaticon](#) and infographics & images by [Freepik](#)