CS 563 KNOWLEDGE DISCOVERY & DATA MINING

DEPARTMENT OF COMPUTER SCIENCE & INFORMATION SYSTEMS

An Analysis on Various Machine Learning Models and Algorithms Using WEKA Software

[What Attributes Contribute Most to Matching Romantically with a Partner?]

PROFESSOR	TEAM MEMBERS	
DR. CHRIS NIKOLOPOULOS		
	IMTIAZ AHMED SHAIK ***(ID: 350607)*** CONNOR STEELE (***ID: 319964)***	**SUBMISSION DATE** 06 MAY 20223 **TERM** **SPRING'2023**

TABLE OF CONTENTS

ABSTRACT	**3**
OBJECTIVE	**3**
PREPROCESSING	**4**
LOADING THE FILE INTO WEKA	**4**
FILLING IN MISSING VALUES	**5,6**
OTHER PREPROCESSING STEPS	**7**
APPLYING MODELS	**8**
POSTPROCESSING	**10**
STACKING	**13**
CONCLUSION	**20**
REFERENCES	**21**
APPENDIX	**22**

ABSTRACT

A speed dating experiment was conducted at the Columbia Business School. Individuals were set on blind dates and asked a series of questions about themselves and their partner. They were asked a series of questions about themselves and their partner. Then they were asked if they would like to see their partner again. The answers to all these questions were recorded and the data was imported into Waikato Environment for Knowledge Analysis (WEKA) software for analysis.

OBJECTIVE

The goal of this project is to use WEKA to analyze this “Speed Dating” dataset. First, WEKA’s built in algorithms will be used for cleansing the data. Next, WEKA will use various models and algorithms to help in the understanding of what attributes contribute most to individuals getting a match, ending with classification as “match” or “no-match”. Attributes range from career, race, and income to attractiveness and intelligence, both of themselves and their potential partner. See Appendix A for the entire list of attributes. After determining what the best models are for the analysis, stacking will be implemented to produce the overall best model. With this improved model, when a new user is asked the same questions in the dataset, the probability they will get a match can be predicted with a certain level of confidence.

PREPROCESSING

- ***Loading the File into WEKA***

Before running any algorithm, the data first had to be cleaned. For example, the data may contain missing values, typos, or unnecessary attributes that may make certain algorithms more prone to error. The first preprocessing step occurred when trying to convert the .csv dataset into the .arff, with the error shown below in Figure 1.

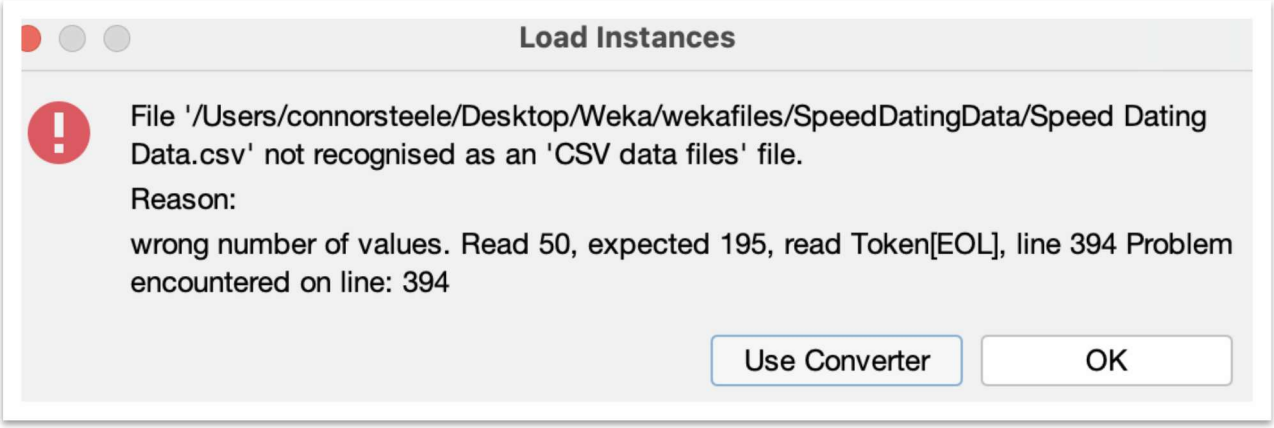

Figure 1. Error Loading the .csv File

After some investigation, it was discovered that the .csv file could not be converted due to the presence of apostrophes in some cells. To solve this issue, the apostrophes were removed from the file using the Excel "find and replace" function. After this, WEKA was able to read and convert the .csv file into a .arff file to run the analysis.

- ***Filling in Missing Values***

A significant number of holes, or missing values, were scattered throughout the dataset. Most holes in the numeric data were handled by using the ReplaceMissingWithUserConstant (RMWUC), ReplaceMissingValues (RMV), or RemoveWithValues (RWV) filters. RMWUC was

used on attributes with a clearly distinct mode, where the mode was used in place of the missing value. An example is shown below in Figure 2.

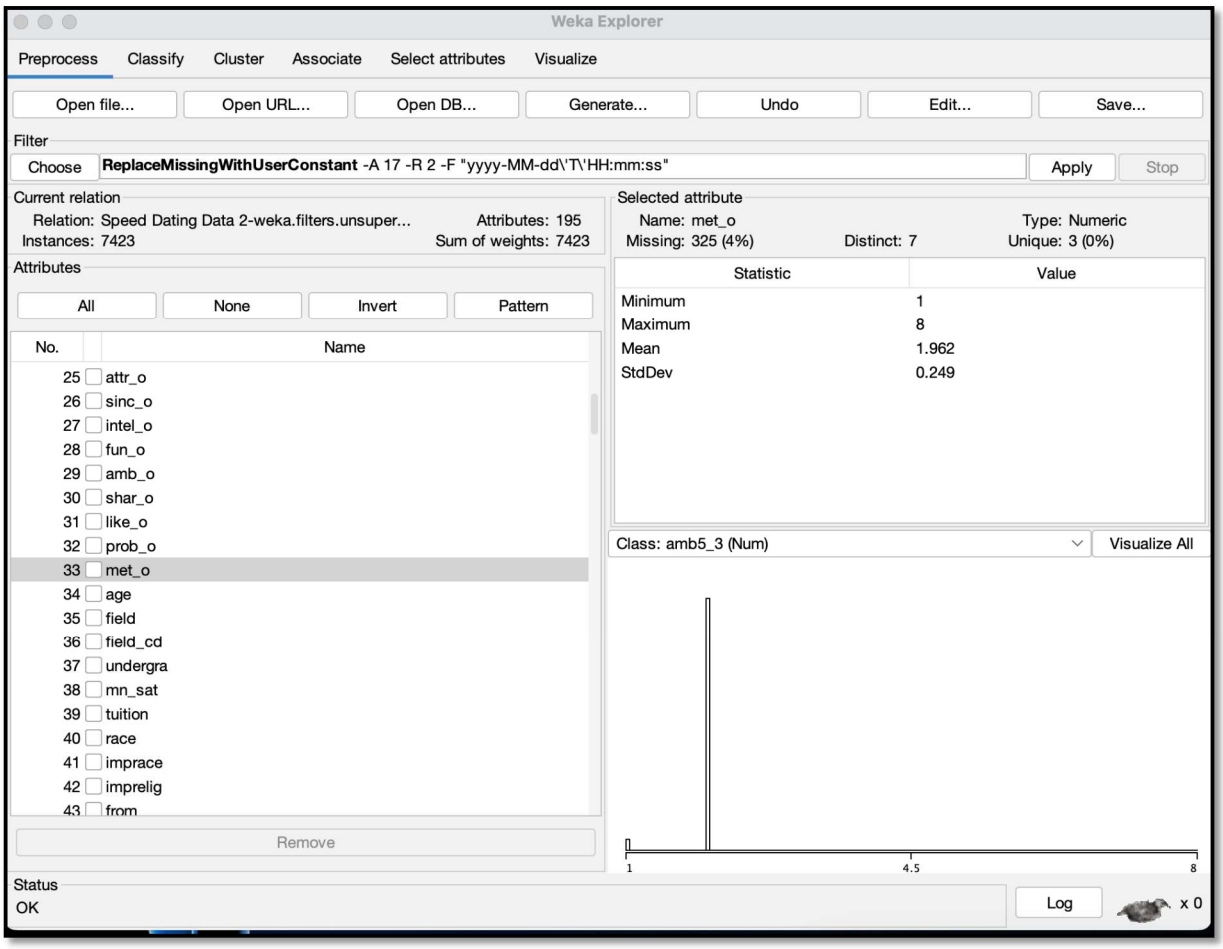

Figure 2. RMWUC for "met_o"

In the case above, as shown in the bar graph, the mode is clearly defined (corresponding to a value of 2). Therefore, the missing values were filled in with the mode.

RWV was used in attributes where a clearly distinct mode was not defined. In other words, the data followed more of a uniform distribution. An example is shown below in Figure 3.

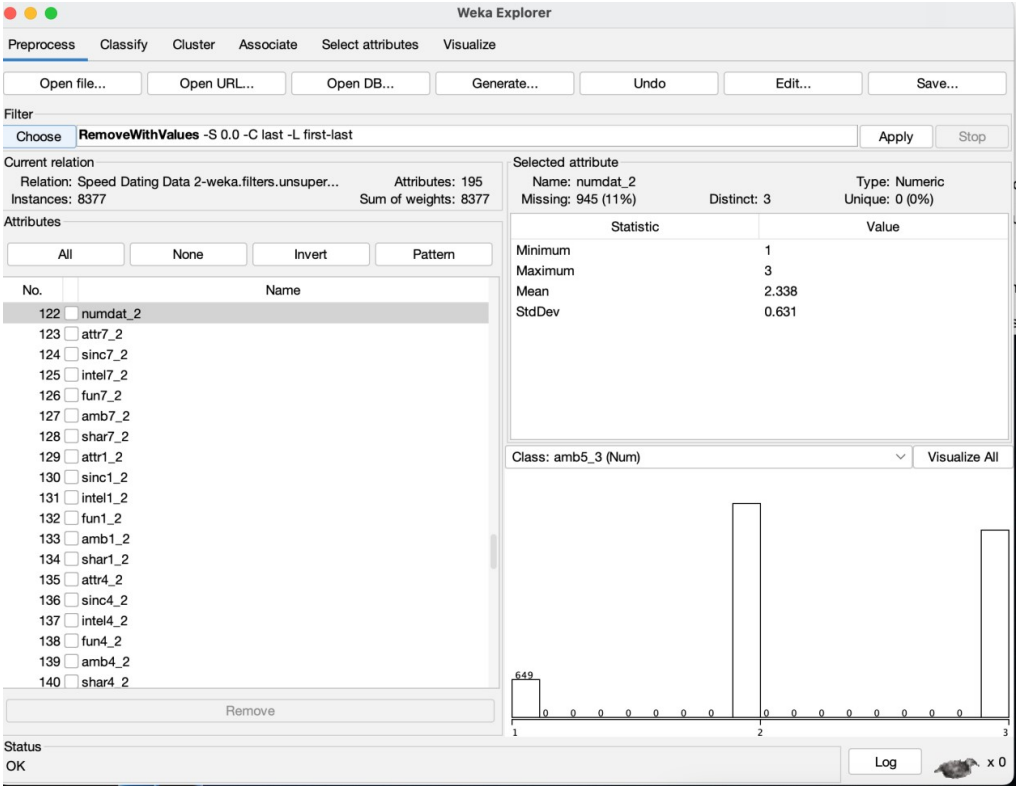

Figure 3. RWV for "numdat_2"

As shown above, there's no clearly defined mode and the distribution follows more of a uniform distribution. Therefore, the instances with the missing values were deleted.

RWV was used with attributes that followed a normal distribution. See Figure 4 below.

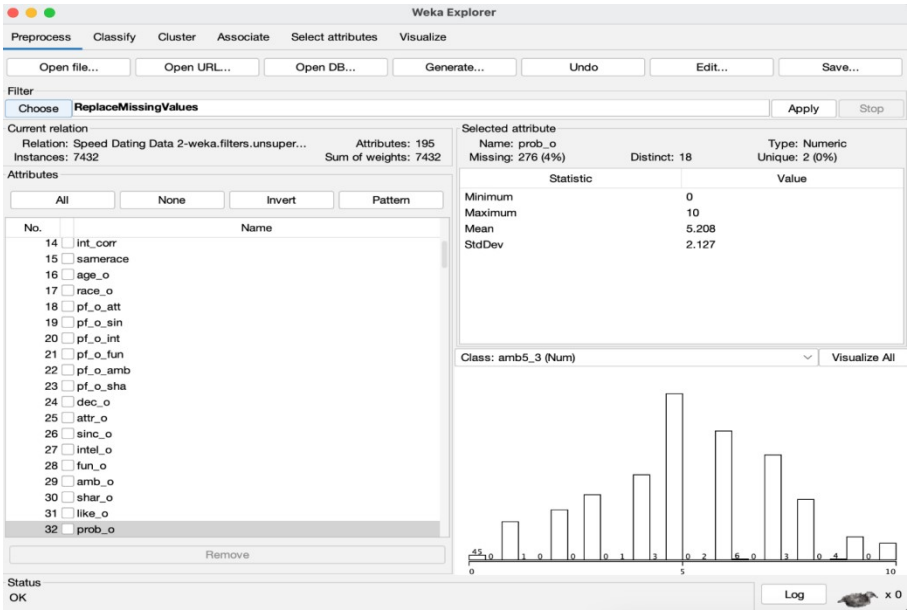

Figure 4. RMV for "prob_o"

From Figure 4 above, the distribution is assumed to be normally distributed. The missing values were replaced with the modes and means using RMW. This was performed at the end of data cleansing because RMW works on the entire dataset.

Other Preprocessing Steps

To remove complexity, judgment was made to delete some attributes altogether. For example, some attributes had the vast majority of their data missing (ex: 70% +). These attributes were deleted. Other attributes were deleted based on the judgment that they would have little impact on matches, like ID number, unique ID number and wave number.

Some questions in the questionnaire would obviously involve a match, leading to no valuable insight. So these had to be deleted as well. For example, some questions were only answered if individuals matched in the first place. See the "followup2/ Time3" section in Appendix A. Similarly, among users who decide "yes" (dec = 1) where their partner also decides "yes" (dec_0 = 1), they will likely match. Therefore, these attributes were deleted.

As for the remaining cleansing procedures, "Match" was changed from numeric to nominal using the NumerictoNominal filter. Then RenameNominalValues was used to change 1 to "match" and 0 to "no_match".

Most nominal data types had numeric versions provided. The nominal versions were deleted.

After the data was cleaned and all missing values were replaced accordingly, the dataset shrunk from **8378 instances and 195 attributes to 3558 instances and 169 attributes.**

APPLYING MODELS

GreedyStepwise + CfsSubset Eval and BestFirst + CfsSubsetEval were then used to select the best attributes. They both suggested the same attributes, shown below in Figure 5.

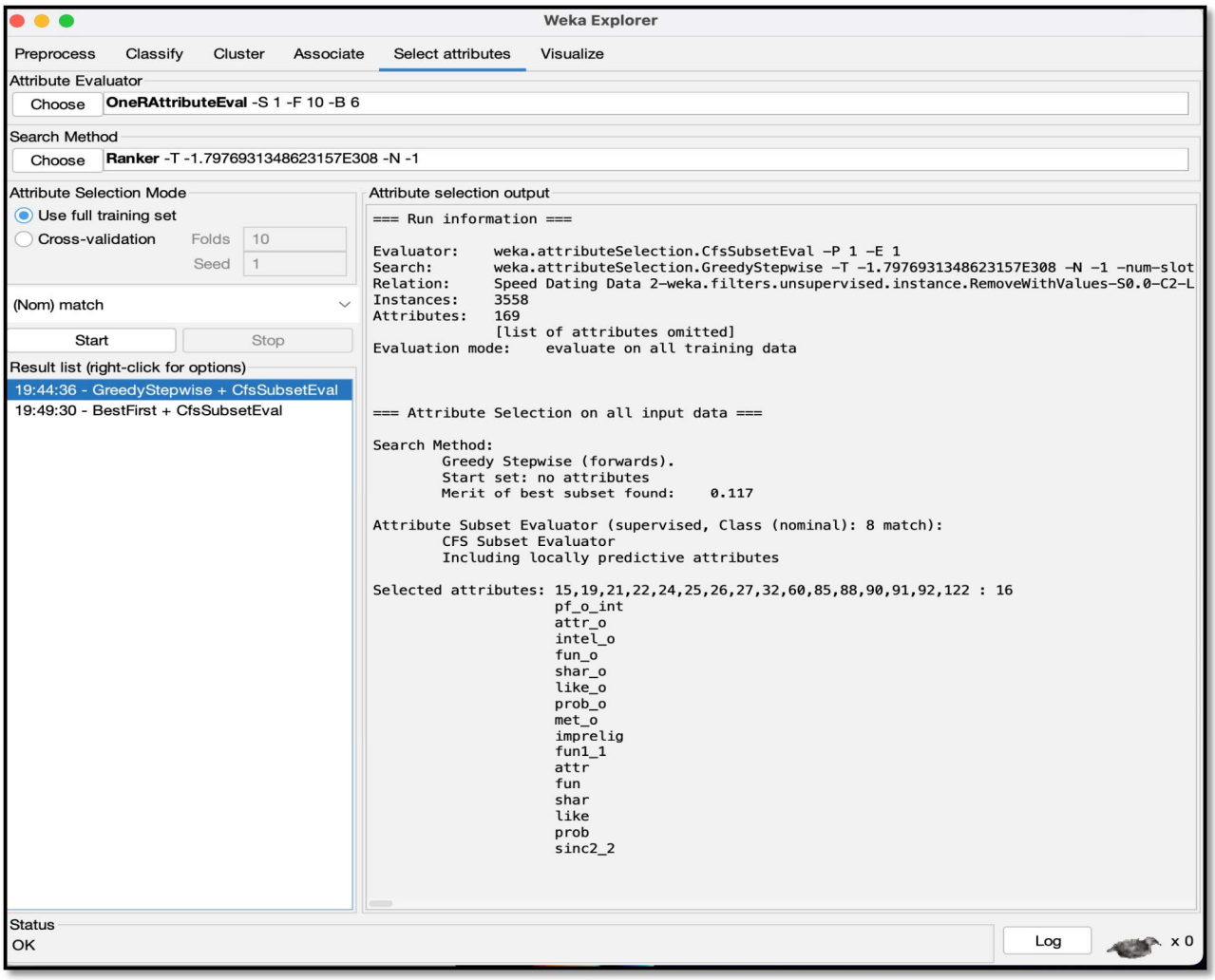

Figure 5. Selected Attributes

Using these attributes and Match, 10 models were run 10 times each using Percentage split at 66% train, and the remainder was tested. The first test for each model was run with a value of 1 for “Random seed for XVal / % Split” . This value increased by 1 for each trial. An example output for tree.RandomForest is shown below with the seed value at 1 in Figure 6.

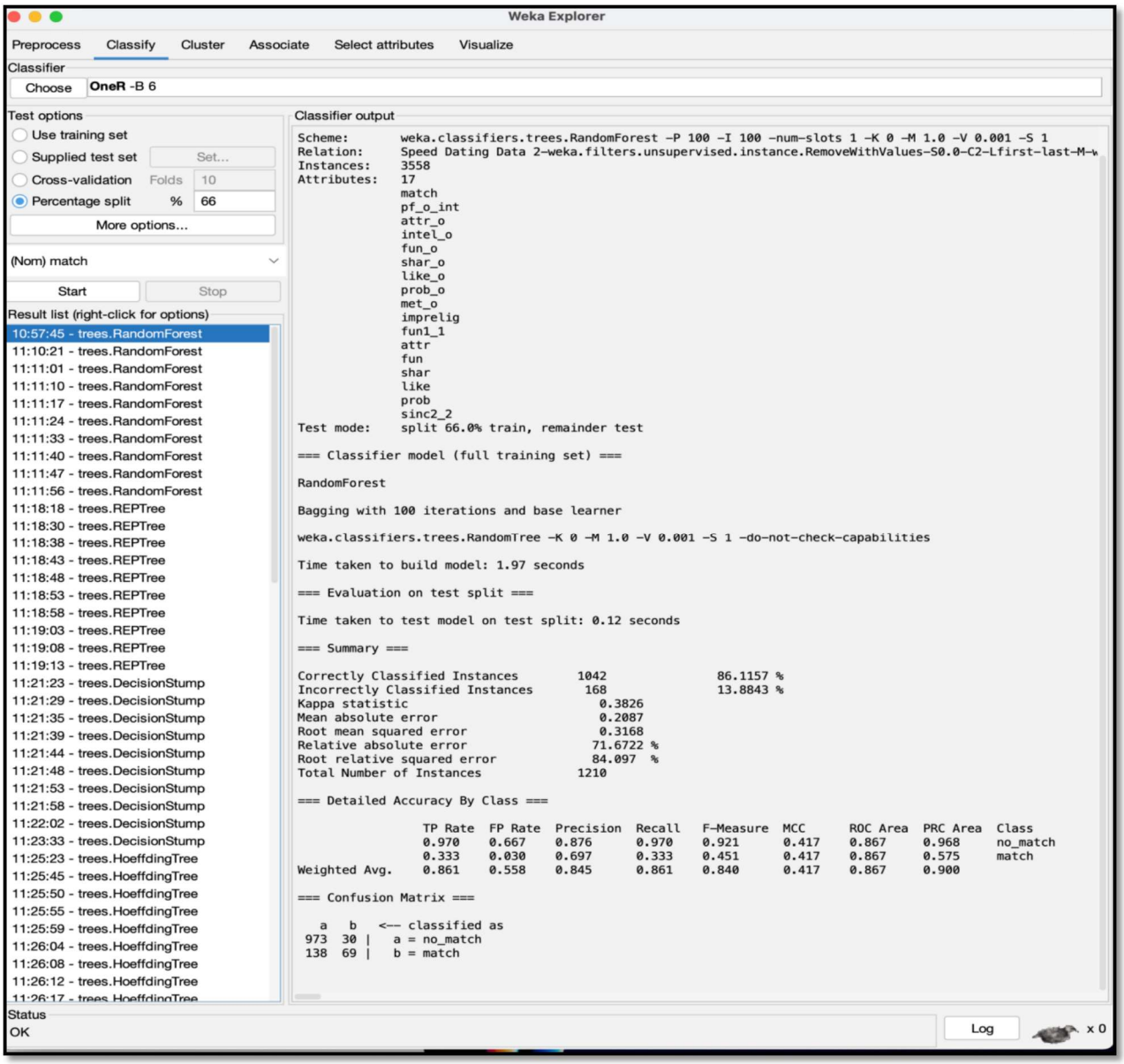

Figure 6. Trial 1 Output for RandomForest

As shown above in Figure 6, the root mean squared error is 0.3168 for the first trial of RandomForest.

The final "cleansed" data set ready for postprocessing had **3558 instances and 17 attributes,** with the top 17 attributes (including "Match") shown above.

POST PROCESSING

A summary of the outputs is shown below in Table 1, ranked from lowest root mean error (top) to highest root mean error (bottom).

Table 1. Comparing Root Mean Squared Error with Various Models

Root Mean Squared Error Per Trial											
Classifier	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**Average**
trees.RandomForest	0.3168	0.3206	0.3227	0.3238	0.3158	0.3297	0.3246	0.32	0.3327	0.3154	**0.32221**
functions.Logistic	0.3219	0.3254	0.3323	0.3332	0.3256	0.3361	0.3311	0.3257	0.3372	0.3213	**0.32898**
meta.Bagging	0.3195	0.3294	0.3328	0.334	0.3259	0.3403	0.3309	0.3253	0.3416	0.3227	**0.33024**
trees.REPTree	0.3548	0.3569	0.3494	0.3573	0.3465	0.3582	0.3617	0.3455	0.3656	0.3476	**0.35435**
trees.DecisionStump	0.361	0.3725	0.3678	0.371	0.3625	0.3844	0.3689	0.3717	0.3777	0.3665	**0.3704**
bayes.NaiveBayesMultinomial	0.381	0.366	0.3856	0.3704	0.3691	0.384	0.3764	0.3759	0.3785	0.3744	**0.37613**
trees.HoeffdingTree	0.3767	0.3829	0.3816	0.3843	0.3761	0.397	0.3836	0.3843	0.3927	0.3823	**0.38415**
misc.InputMappedClassifier	0.3767	0.3829	0.3816	0.3843	0.3761	0.397	0.3836	0.3843	0.3927	0.3823	**0.38415**
rules.OneR	0.4176	0.4166	0.4379	0.4245	0.4126	0.4388	0.4225	0.4215	0.435	0.4254	**0.42524**
lazy.Ibk	0.4482	0.4358	0.4625	0.4479	0.4389	0.4525	0.4339	0.4445	0.4507	0.4398	**0.44547**

As shown above in Table 1, in terms of the average testing error, the top four classifiers are RandomForest, Logistic, Bagging, and REPTree. A “tournament-style” ranking system was used on these four classifiers. Of these four, the best and worst were paired together for a paired t-test, and the middle two were then paired for a t-test. The two winners of this round were then paired for the second round, determining the overall winner. See Tables 2 and 3 below for the results.

Table 2. Comparing Top Four Models

1st round	
trees.RandomForest avg error	0.32221
trees.REPTree avg error	0.35435
md	-0.03214
k	10
var	0.00025824
Q	-6.3245553
df	9
Confidence Level	.95
alpha	0.025
t from table	2.262
interval	[-2.262,2.262]
Better model	**REPTree**
functions.Logistic avg error	0.32898
meta.Bagging avg error	0.33024
md	-0.00126
k	10
var	3.969E-07
Q	-6.3245553
df	9
Confidence Level	0.95
alpha	0.025
t from table	2.262
interval	[-2.262,2.262]
Better model	**Bagging**

Table 3. 2nd Round

2nd round	
REPTree avg error	0.35435
Bagging avg error	0.33024
md	0.02411
k	10
var	0.00014532
Q	6.32455532
df	9
Confidence Level	95%
alpha	0.025
t from table	2.262
interval	[-2.262,2.262]
Better model	**REPTree**

Where,
K = number of test trials
md = difference in avg error
var = variance
Q = md/(sqrt(var/k))
df = degrees of freedom

As shown in Table 2 above, with 95% confidence, the best two models are REPTree and bagging. For the second round, between these two models, REPTree was the overall best model with 95% confidence as shown in Table 3 above.

STACKING

Stacking was then implemented to find an even more accurate model using the Weka Experiment Environment. The “Match” attribute had to be manually selected as the class in Weka Explorer. First, the same 10 classifiers were used in the stacking. The first trial was used with RandomForest as the metaClassifier, as this originally had the lowest root mean error. See Figure 7 below for the setup and Figure 8 for the output with root mean square error.

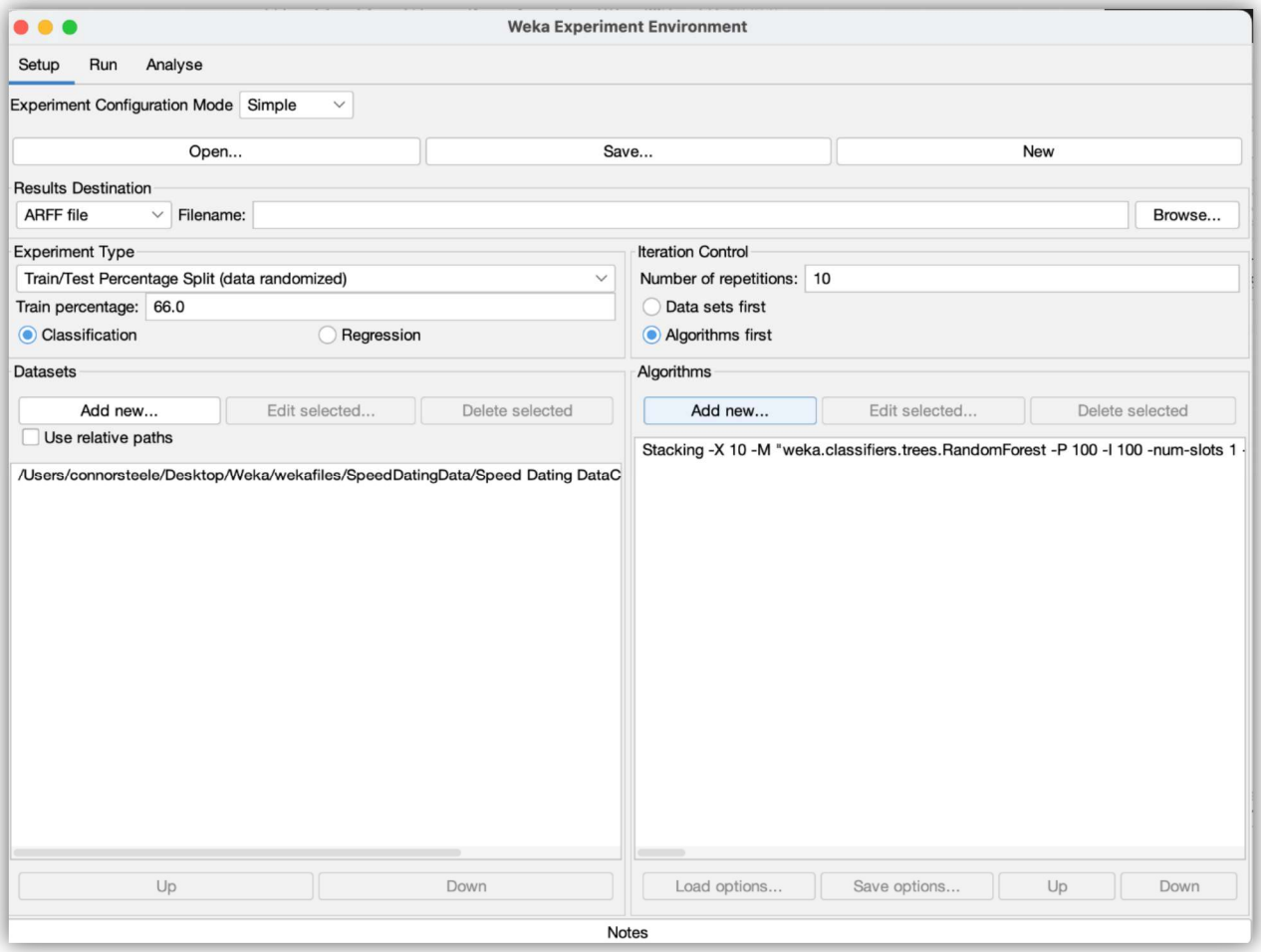

Figure 7. Stacking Setup Page Configuration

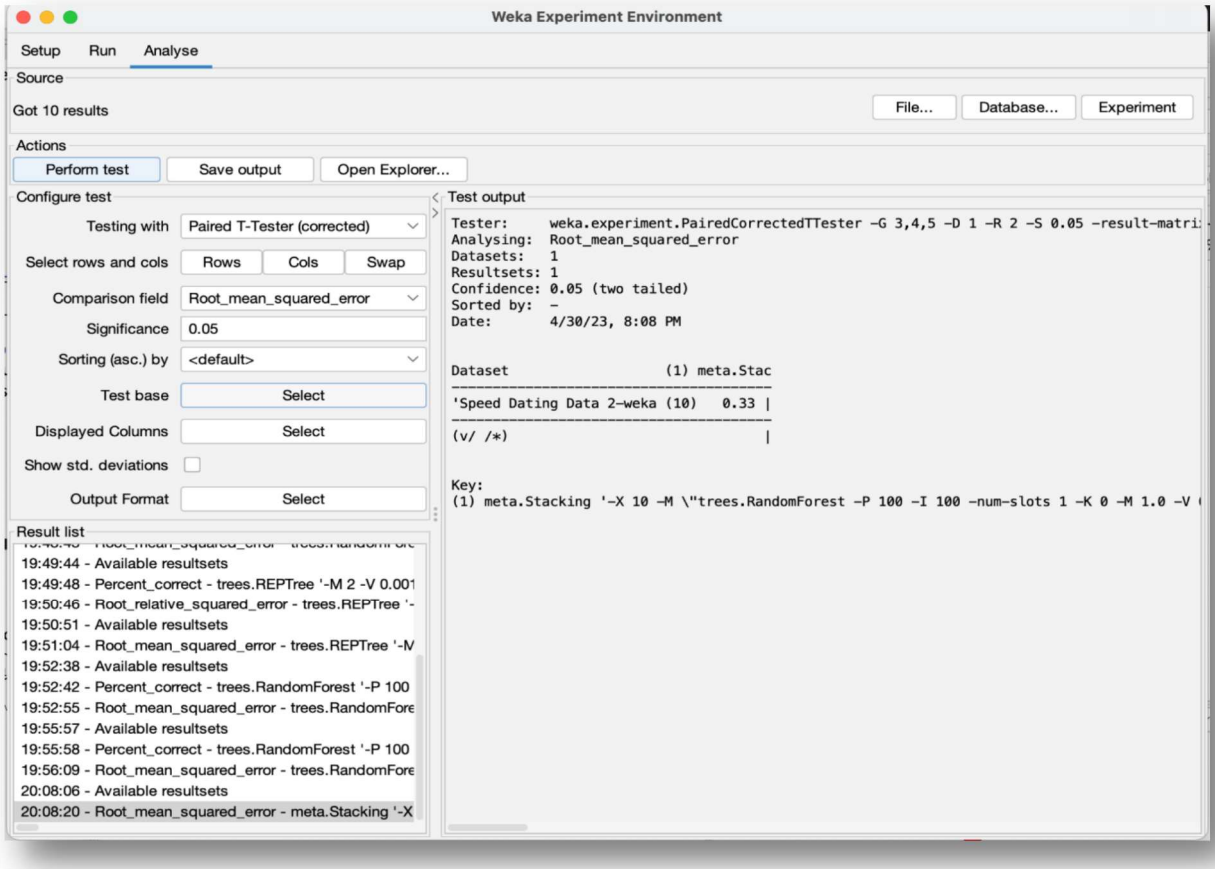

Figure 8. Stacking Test Output With All 10 Classifiers Using RandomForest metaClassifier

As shown in Figure 8 above, the root mean squared error is 0.33. In terms of root mean squared error alone, this is better than most of the other classifiers.

Another test using stacking was run using only the top four classifiers depicted in Table 1 with RandomForest metaClassifier again. This also yielded a root mean squared error of 0.33. Since this new stacking model contained less classifiers, by Occam's razor principle, this algorithm was selected to have a t test compared with REPTree from above to determine the overall "winner". The output is shown below in Figure 9.

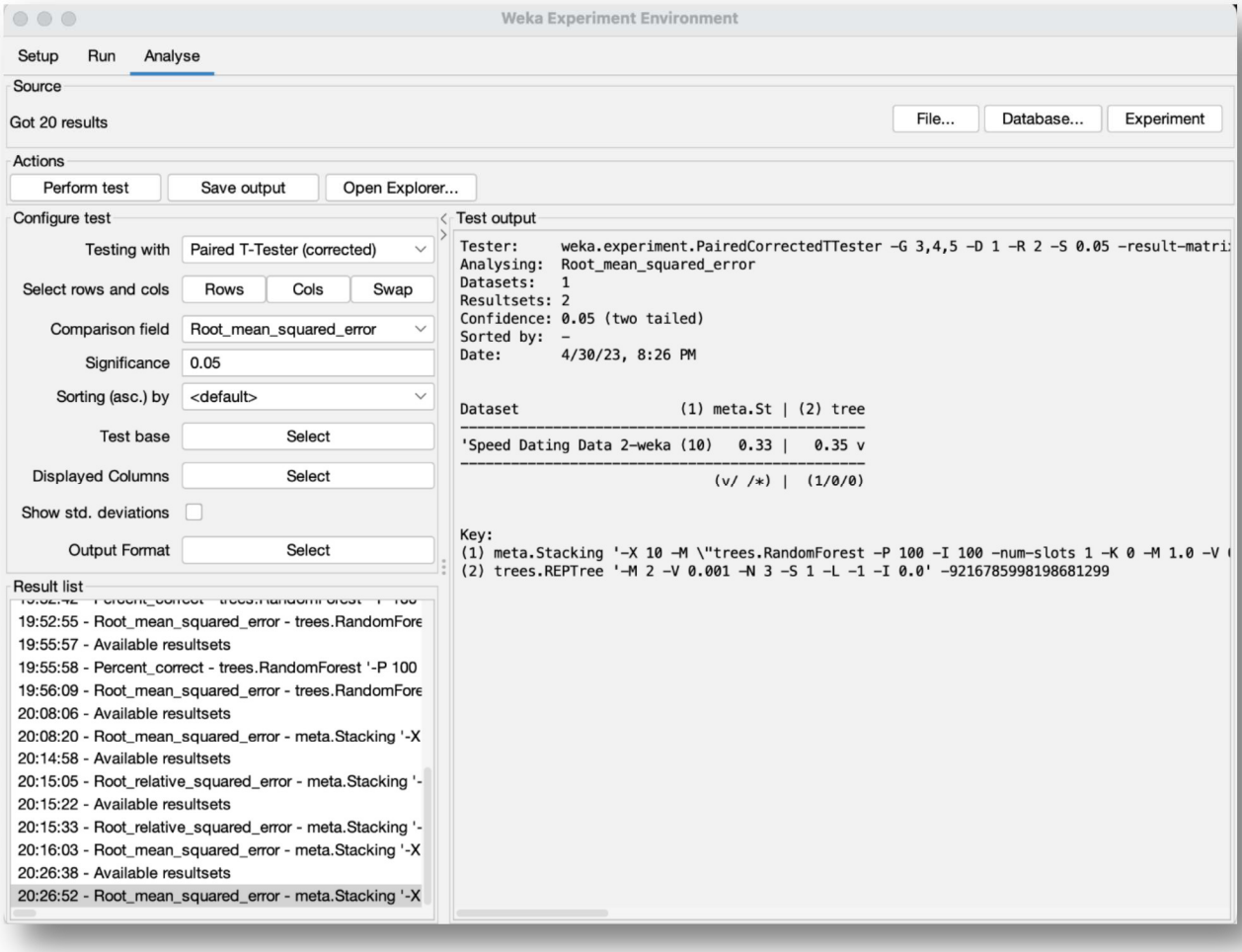

Figure 9. Paired t-test with Stacking Classifier and REPTree

As shown above, when the stacking classifier with the top four classifiers in Table 1 is paired with the previous "winner" REPTree, REPTree is better with 95% confidence. However, when running other tests such as percentage correct and number correct, the stacking classifier did perform better, but neither was statistically significant (i.e, neither had any "*" or "v" symbols next to the outputs). See Figures 10 and 11 below.

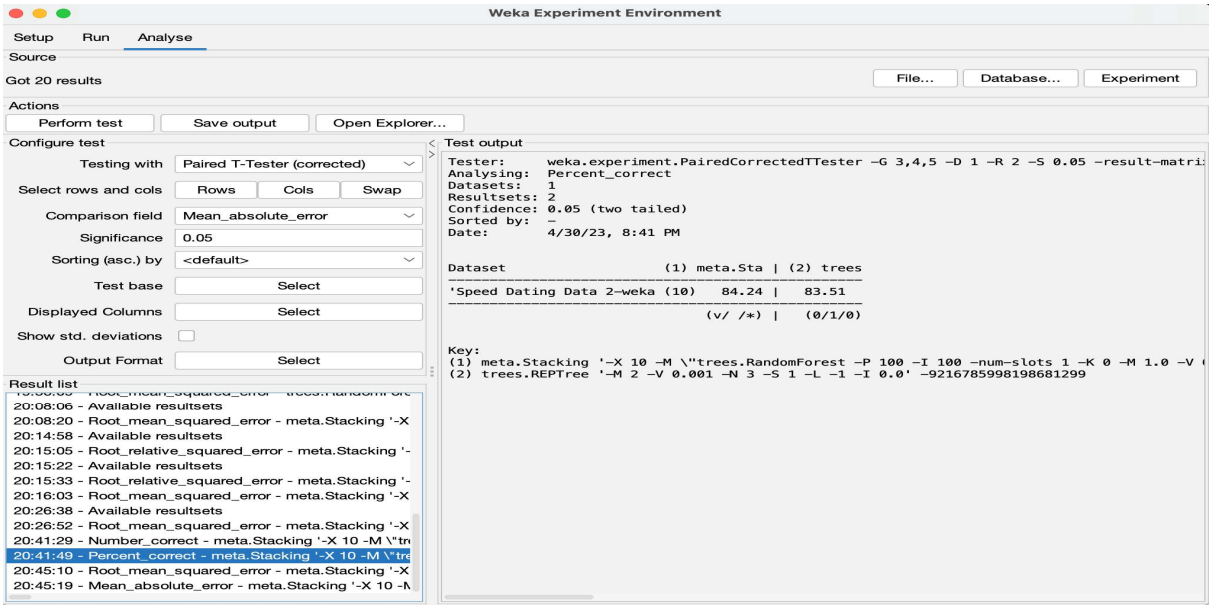

Figure 10. No Conclusion For Percent Correct Between Stacking and REPTree Paired t-test

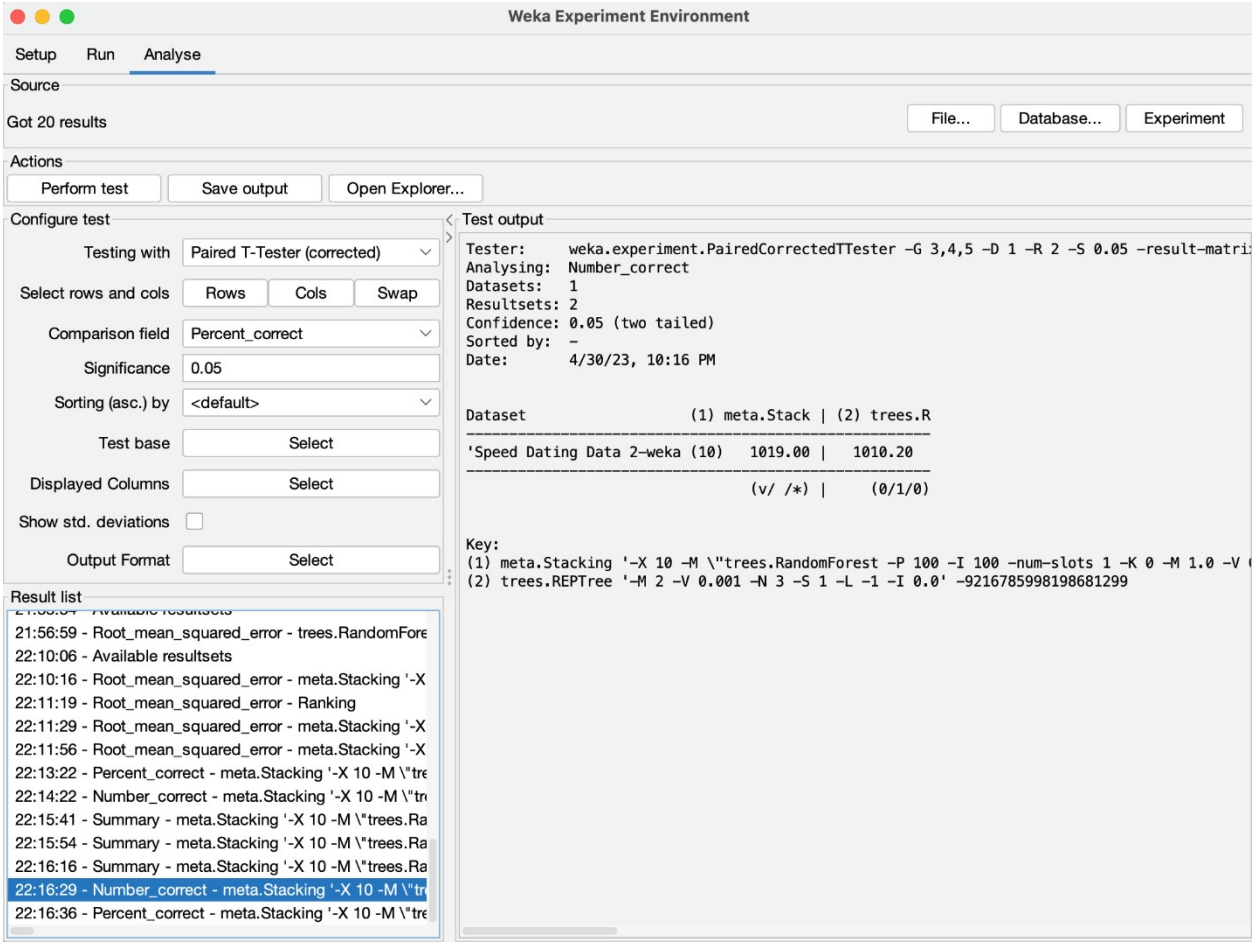

Figure 11. No Conclusion For Number Correct Between Stacking and REPTree Paired t-test

A summary of the results for all the t-tests is shown below in Figure 12 for visualization purposes.

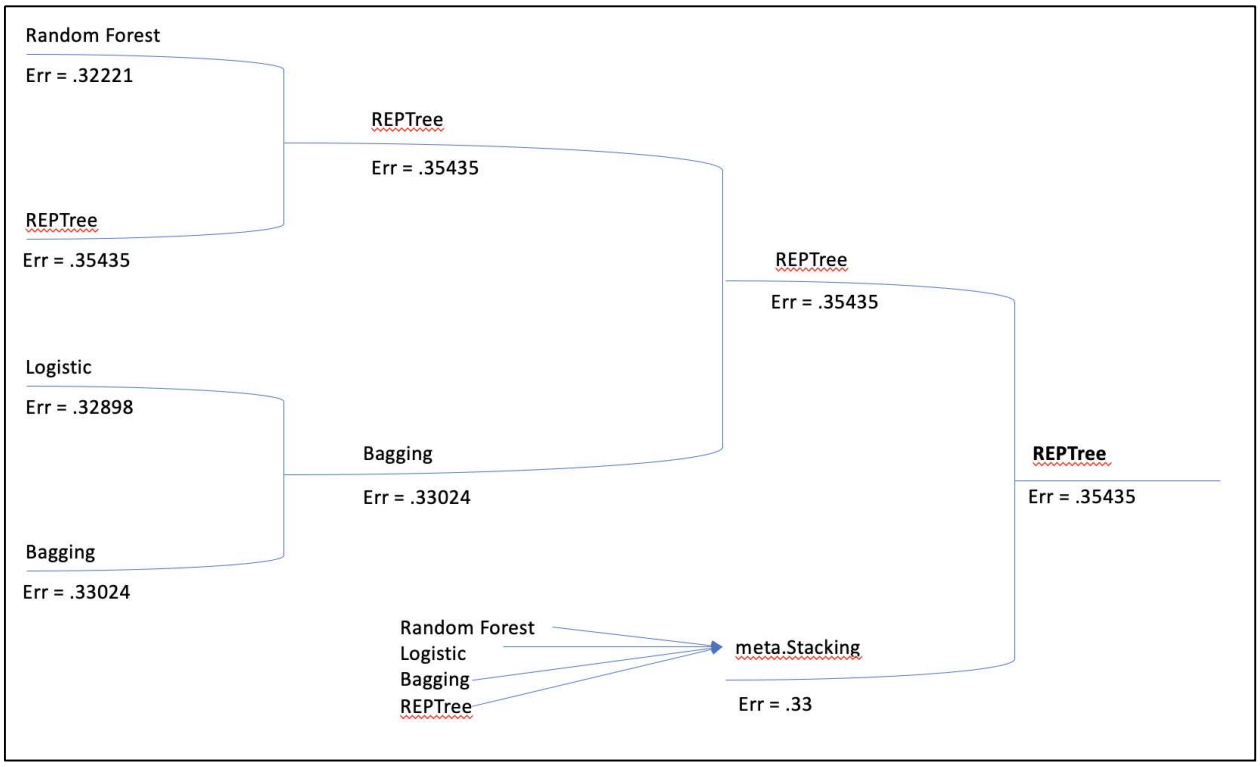

Figure 12. Paired t-test Results Using Root Mean Squared Error at 95% Confidence

The output for the "winning" classifier, REPTree, is shown below in Figure 13. The tree visualizer is in Figure 14.

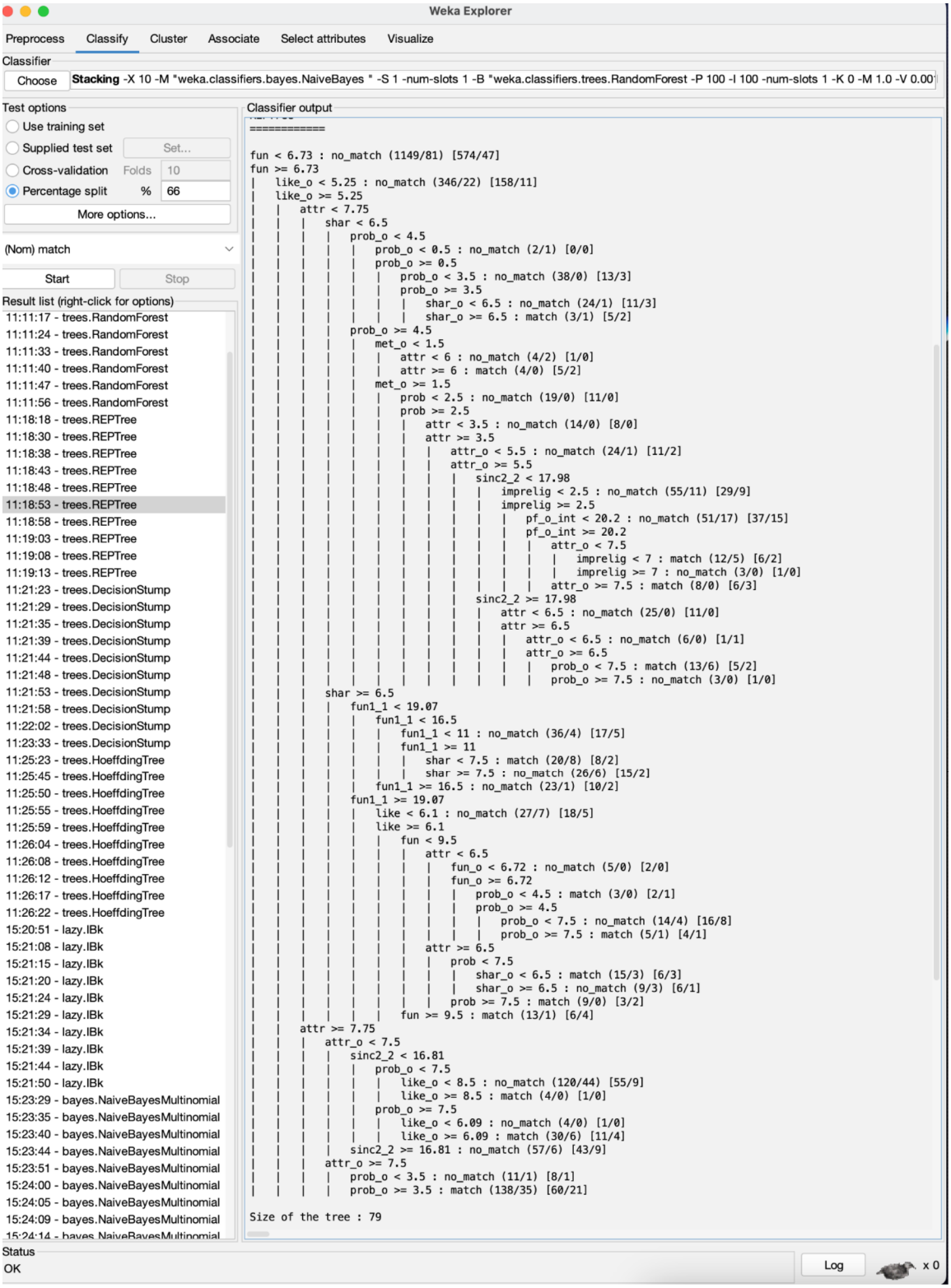

Figure 13. "Winner" REPTree Output (Trial 6)

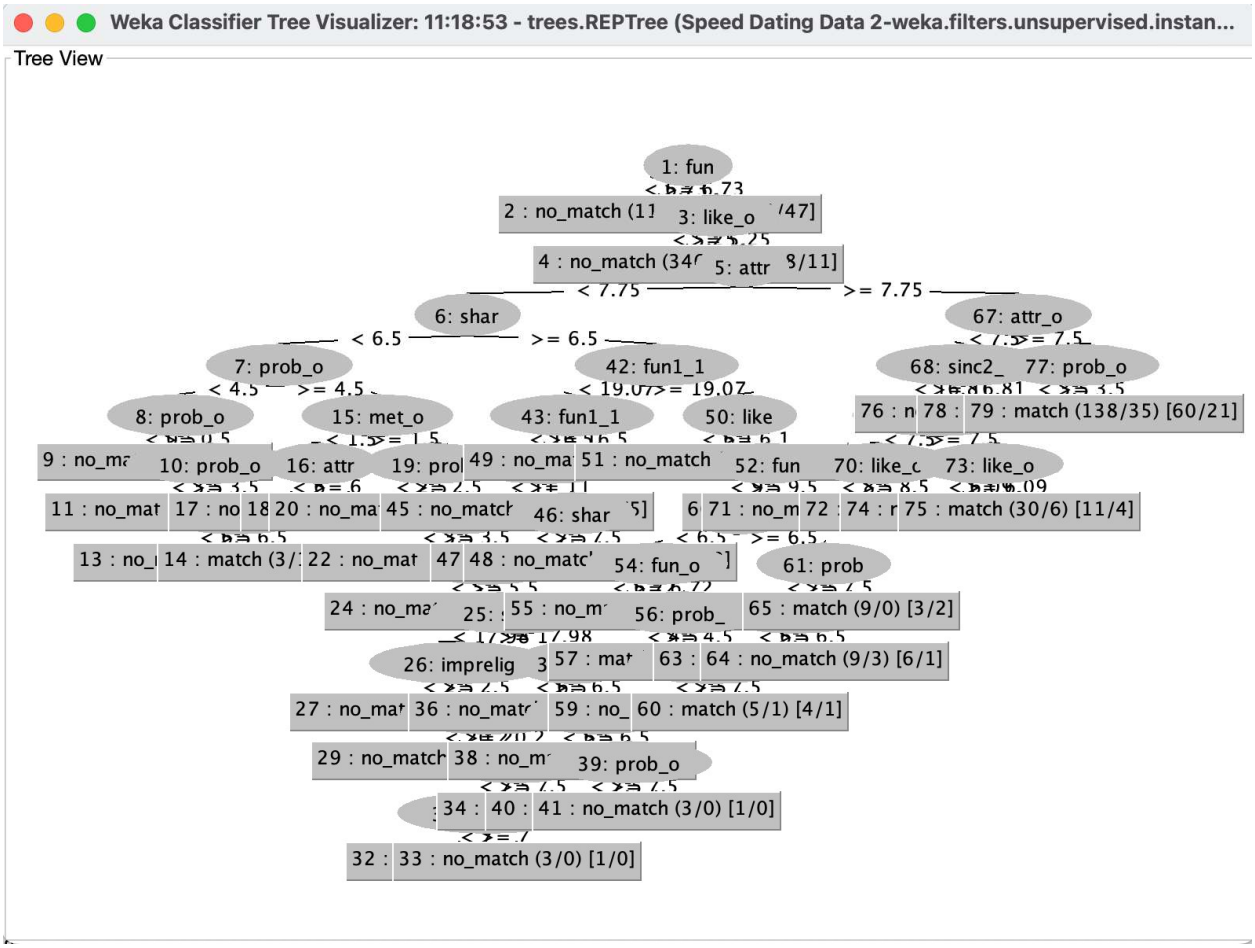

Figure 14. "Winner" REPTree Visualizer (Trial 6)

For the first few leaves, the first attribute likely to lead to a match is for the individual to rate their date at least a 6.73/10 on being fun (see Figure 13 at the top). Then, they have to like this person at least 5.25/10. The decision tree continues as shown above. See Appendix A for the description of the attributes.

CONCLUSION

In conclusion the four classifiers with the lowest average root mean squared error were RandomForest, Logistic, Bagging, and REPTree. However when the t-test was performed among these four, REPTree was the overall winner with 95% confidence, despite RandomForest having the lowest average root mean squared error. Then stacking was performed and compared with REPTree using another t-test. When comparing root mean squared error, REPTree was significantly better at 95% confidence even though stacking had more instances correctly classified and lower average root mean square error.

REFERENCES

- "Demonstrating Stacking in Weka" *YouTube*, uploaded by WIT Solapur - Professional Learning Community, 29 April, 2021, https://www.youtube.com/watch?v=P4uaMbWPYN0

- "How to replace missing values in weka using different methods | Weka Tutorial" *YouTube, uploaded* by DWBIADDA VIDEOS, 10 December, 2020, https://www.youtube.com/watch?v=P3mN9g-hvmk
- "(Solved - Wrong number of values) WEKA Tool Dataset Error | read Token[EOL] | RK" *YouTube,* uploaded by RK Keynotes, 15 December, 2020, https://www.youtube.com/watch?v=xRWKaZuhu7E
- "Weka Tutorial 19: Outliers and Extreme Values (Data Preprocessing)" *YouTube,* uploaded by Rushdi Shams, 1 November, 2021, https://www.youtube.com/watch?v=WrjpO7CmUoQ&t=119s
- "Weka Tutorial 24: Model Comparison (Model Evaluation)" *YouTube,* uploaded by Rushdi Shams, 20 August, 2023, https://www.youtube.com/watch?v=WqEAD_Z6h0E
- Witten I. and Frank E., DATA MINING: Practical Machine Learning Tools and Techniques, Morgan Kaufmann Publishers, 2020.

APPENDIX

Appendix A.
Speed Digital Data Key

iid: unique subject number, group(wave id gender)

id: subject number within wave

gender: Female=0
Male=1

idg: subject number within gender, group(id gender)

condtn:
1=limited choice
2=extensive choice
wave:

Wave #	Date	Preference Scale	Variations	# Males	# Females
1	October 16th '02	100 pt alloc.		10	10
2	October 23rd '02	100 pt alloc.		16	19
3	November 12th '02	100 pt alloc.		10	9
4	November 12th '02	100 pt alloc.		18	18
5	November 20th, '02	100 pt alloc.	undergrads	10	10
6	March 26th '03	1-10 scale		5	5
7	March 26th '03	1-10 scale		16	16
8	April 2nd '03	1-10 scale		10	10
9	April 2nd '03	1-10 scale		20	20
10	September 24th '03	100 pt alloc.		9	9
11	September 24th '03	100 pt alloc.		21	21

12	October 7th '03	100 pt alloc.	Budget: only allowed to yes yes to 50% of the people that met	14	15
13	October 8th '03	100 pt alloc.	Different M.C.	9	10
14	October 8th '03	100 pt alloc.	Different M.C.	18	20
15	February 24th '04	100 pt alloc.		19	18
16	February 25th '04	100 pt alloc.		8	6
17	February 25th '04	100 pt alloc.		14	10
18	April 6th '04	100 pt alloc.	brought a magazine	6	6
19	April 6th '04	100 pt alloc.	brought a book	15	16
20	April 7th '04	100 pt alloc.	brought a book	8	6
21	April 7th '04	100 pt alloc.	brought a magazine	22	22

round: number of people that met in wave

position: station number where met partner

positin1: station number where started

order: the number of date that night when met partner

partner: partner's id number the night of event

pid: partner's iid number

match 1=yes, 0=no

int_corr: correlation between participant's and partner's ratings of interests in Time 1

samerace: participant and the partner were the same race. 1= yes, 0=no

age_o: age of partner

race_o: race of partner

pf_o_att: partner's stated preference at Time 1 (attr1_1) for all 6 attributes

dec_o: decision of partner the night of event

attr_o: rating by partner the night of the event, for all 6 attributes

signup/Time1:
[Survey filled out by students that are interested in participating in order to register for the event.]
age:

field: field of study

field_cd: field coded
1= Law
2= Math
3= Social Science, Psychologist
4= Medical Science, Pharmaceuticals, and Bio Tech
5= Engineering
6= English/Creative Writing/ Journalism
7= History/Religion/Philosophy
8= Business/Econ/Finance
9= Education, Academia
10= Biological Sciences/Chemistry/Physics
11= Social Work
12= Undergrad/undecided
13=Political Science/International Affairs
14=Film
15=Fine Arts/Arts Administration
16=Languages
17=Architecture
18=Other

undergrd: school attended for undergraduate degree

mn_sat: Median SAT score for the undergraduate institution where attended. Taken from Barron's 25th Edition college profile book. Proxy for intelligence.

tuition: Tuition listed for each response to undergrad in Barron's 25th Edition college profile book.

race:
Black/African American=1

European/Caucasian-American=2
Latino/Hispanic American=3
Asian/Pacific Islander/Asian-American=4
Native American=5
Other=6

imprace:
How important is it to you (on a scale of 1-10) that a person you date be of the same racial/ethnic background?

imprelig:
How important is it to you (on a scale of 1-10) that a person you date be of the same religious background?

from:
Where are you from originally (before coming to Columbia)?

zipcode:
What was the zip code of the area where you grew up?

income:
Median household income based on zipcode using the Census Bureau website:
http://venus.census.gov/cdrom/lookup/CMD=LIST/DB=C90STF3B/LEV=ZIP
When there is no income it means that they are either from abroad or did not enter their zip code.

goal:
What is your primary goal in participating in this event?
Seemed like a fun night out=1
To meet new people=2
To get a date=3
Looking for a serious relationship=4
To say I did it=5
Other=6

date:
In general, how frequently do you go on dates?
Several times a week=1
Twice a week=2
Once a week=3
Twice a month=4
Once a month=5
Several times a year=6
Almost never=7

go out:
How often do you go out (not necessarily on dates)?

Several times a week=1
Twice a week=2
Once a week=3
Twice a month=4
Once a month=5
Several times a year=6
Almost never=7

career:
What is your intended career?

career_c: career coded
1= Lawyer
2= Academic/Research
3= Psychologist
4= Doctor/Medicine
5=Engineer
6= Creative Arts/Entertainment
7= Banking/Consulting/Finance/Marketing/Business/CEO/Entrepreneur/Admin
8= Real Estate
9= International/Humanitarian Affairs
10= Undecided
11=Social Work
12=Speech Pathology
13=Politics
14=Pro sports/Athletics
15=Other
16=Journalism
17=Architecture

12. How interested are you in the following activities, on a scale of 1-10?
sports: Playing sports/ athletics
tvsports: Watching sports
excersice: Body building/exercising
dining: Dining out
museums: Museums/galleries
art: Art
hiking: Hiking/camping
gaming: Gaming
clubbing: Dancing/clubbing
reading: Reading
tv: Watching TV
theater: Theater
movies: Movies
concerts: Going to concerts
music: Music

shopping: Shopping
yoga: Yoga/meditation

exphappy:
Overall, on a scale of 1-10, how happy do you expect to be with the people you meet during the speed-dating event?

expnum:
Out of the 20 people you will meet, how many do you expect will be interested in dating you?

We want to know what you look for in the opposite sex.
Waves 6-9: Please rate the importance of the following attributes in a potential date on a scale of 1-10 (1=not at all important, 10=extremely important):
Waves 1-5, 10-21: You have 100 points to distribute among the following attributes -- give more points to those attributes that are more important in a potential date, and fewer points to those attributes that are less important in a potential date. Total points must equal 100.
attr1_1
Attractive
sinc1_1
Sincere
intel1_1
Intelligent
fun1_1
Fun
amb1_1
Ambitious
shar1_1
Has shared interests/hobbies

Now we want to know what you think MOST of your fellow men/women look for in the opposite sex.
Waves 6-9: Please rate the importance of the following attributes on a scale of 1-10 (1=not at all important, 10=extremely important):
Waves 10-21 : You have 100 points to distribute among the following attributes -- give more points to those attributes that you think your fellow men/women find more important in a potential date and fewer points to those attributes that they find less important in a potential date. Total points must equal 100.

attr4_1
Attractive
sinc4_1
Sincere
intel4_1
Intelligent
fun4_1
Fun
amb4_1

Ambitious
shar4_1
Shared Interests/Hobbies

What do you think the opposite sex looks for in a date?
Waves 6-9: Please rate the importance of the following attributes on a scale of 1-10 (1=not at all important, 10=extremely important):
Waves 1-5 and 10-21: Please distribute 100 points among the following attributes -- give more points to those attributes that you think are more important to members of the opposite sex when they are deciding whether to date someone. Total points must equal 100.

attr2_1
Attractive
sinc2_1
Sincere
int2_1
Intelligent
fun2_1
Fun
amb2_1
Ambitious
shar2_1
Has shared interests/hobbies

How do you think you measure up?
Please rate your opinion of your own attributes, on a scale of 1-10 (be honest!):
attr3_1
Attractive
sinc3_1
Sincere
int3_1
Intelligent
fun3_1
Fun
amb3_1
Ambitious

And finally, how do you think others perceive you?
Please rate yourself how you think others would rate you on each of the following attributes, on a scale of 1-10 (1=awful, 10=great)
attr5_1
Attractive
sinc5_1
Sincere
int5_1

Intelligent
fun5_1
Fun
amb5_1
Ambitious
Scorecard:
[Filled out by subjects after each "date" during the event.]

SCORECARD
YOUR ID NUMBER:
Circle “Yes” or “No” below the ID number of each person you meet to indicate whether or not you would like to see him or her again. Rate their attributes on a scale of 1-10: (1=awful, 10=great). If you haven’t formed an opinion based on your conversation, fill in N/A, but please fill in all boxes. This will be TOTALLY confidential and will NOT be shared with anyone. Then, answer the remaining questions for each person you meet.

ID #:	1	3	4	5	6	7	8	9	10
dec									
Decision	1=yes 0=no								
Attributes (1=awful, 10=great)									
Attractive	attr								
Sincere	sinc								
Intelligent	intel								
Fun	fun								
Ambitious	amb								
Shared Interests/Hobbies	shar								
Overall, how much do you like this person? (1=don't like at all, 10=like a lot)	like								

How probable do you think it is that this person will say 'yes' for you? (1=not probable, 10=extremely probable)	prob								
Have you met this person before?	met 1=yes 2=no	yes no	yes no	yes no	yes no	yes no	yes no	yes no	yes no

ID #:	11	12	13	14	15	16	17	18	19	20
Decision	yes no	yes no	yes no	yes no	yes no	yes no	yes no	yes no	yes no	yes no
Attributes (1=awful, 10=great)										
Attractive										
Sincere										
Intelligent										
Fun										
Ambitious										
Shared Interests/Hobbies										
Overall, how much do you like this person? (1=don't like at all, 10=like a lot)										

How probable do you think it is that this person will say 'yes' for you? (1=not probable, 10=extremely probable)										
Have you met this person before?	yes no	yes no	yes no	yes no	yes no	yes no	yes no	yes no	yes no	yes no

match_es:
How many matches do you estimate you will get (a match occurs when you and your partner both check "Yes" next to decision)?: _____

Half way through meeting all potential dates during the night of the event on their scorecard:

Hold up! Now that you are half way through your Speed Dates, we have a few questions for you…

We want to know what you look for in the opposite sex.
Please rate the importance of the following attributes in a potential date on a scale of 1-10: (1=not at all important, 10=extremely important).
attr1_s
Attractive_____
sinc1_s
Sincere _____
intel1_s
Intelligent _____
fun1_s
Fun _____
amb1_s
Ambitious _____
shar1_s
Shared Interests/Hobbies _____

Please rate your opinion of your own attributes, on a scale of 1-10 (1=awful, 10=great) --Be honest!
attr3_s
Attractive_____

sinc3_s
Sincere _____
intel3_s
Intelligent _____
fun3_s
Fun _____
amb3_s
Ambitious _____

followup/Time2:
[Survey is filled out the day after participating in the event. Subjects must have submitted this in order to be sent their matches.]
satis_2:
Overall, how satisfied were you with the people you met? (1=not at all satisfied, 10=extremely satisfied)

length:
Four minutes is:
Too little=1
Too much=2
Just Right=3

numdat_2:
The number of Speed "Dates" you had was:
Too few=1
Too many=2
Just right=3

Now, think back to your yes/no decisions during the Speed Dating event. Try to distribute the 100 points among these six attributes in the way that best reflects the actual importance of these attributes in your decisions. Give more points to those attributes that were more important in your decisions, and fewer points to those attributes that were less important in your decisions. Total points must equal 100.
attr7_2
Attractive
sinc7_2
Sincere
intel7_2
Intelligent
fun7_2
Fun
amb7_2
Ambitious
shar7_2
Has shared interests/hobbies

We want to know what you look for in the opposite sex.
Waves 1-5 and 10-21: You have 100 points to distribute among the following attributes -- give more points to those attributes that are more important in a potential date, and fewer points to those attributes that are less important in a potential date. *Total points must equal 100.*
Waves 6-9: Please rate the importance of the following attributes in a potential date on a scale of 1-10 (1=not at all important, 10=extremely important):
attr1_2
Attractive
sinc1_2
Sincere
intel1_2
Intelligent
fun1_2
Fun
amb1_2
Ambitious
shar1_2
Has shared interests/hobbies

What do you think MOST of your fellow men/women look for in the opposite sex?
You have 100 points to distribute among the following attributes -- give more points to those attributes that you think your fellow men/women find more important in a potential date, and fewer points to those attributes that they find less important in a potential date.
Total points must equal 100.
attr4_2
Attractive
sinc4_2
Sincere
intel4_2
Intelligent
fun4_2
Fun
amb4_2
Ambitious
shar4_2
Shared Interests/Hobbies

What do you think the opposite sex looks for in a date?
Please distribute 100 points among the following attributes -- give more points to those attributes that you think are moreimportant to members of the opposite sex when they are deciding whether to date someone. *Total points must equal 100.*
attr2_2
Attractive
sinc2_2
Sincere
intel2_2

Intelligent
fun2_2
Fun
amb2_2
Ambitious
shar2_2
Has shared interests/hobbies

How do you think you measure up?
Please rate your opinion of your own attributes, on a scale of 1-10 (1= awful and 10=great). Be honest!
attr3_2
Attractive
sinc3_2
Sincere
int3_2
Intelligent
fun3_2
Fun
amb3_2
Ambitious

And finally, how do you think others perceive you?
Please rate yourself how you think others would rate you on each of the following attributes, on a scale of 1-10 (1=awful, 10=great)
attr5_2
Attractive
sinc5_2
Sincere
int5_2
Intelligent
fun5_2
Fun
amb5_2
Ambitious

followup2/ Time3:
[Subjects filled out 3-4 weeks after they had been sent their matches]

SINCE HURRYDATING…

1. Of the matches that you received:
you_call:
(a) How many have you contacted to set up a date?
them_cal:

(b) How many have contacted you?

date_3:
Have you been on a date with any of your matches?
Yes=1
No=2

If you have been on at least one date, please answer the following:
numdat_3:
(a) How many of your matches have you been on a date with so far?
num_in_3
If yes, how many?

What do you look for in the opposite sex?
Please distribute 100 points among the following attributes -- give more to attributes that were more important in your decisions when Hurrydating, and less to attributes that were less important. Total points must equal 100.
We want to know what you look for in the opposite sex.
Please rate the importance of the following attributes in a potential date on a scale of 1-10 (1=not at all important, 10=extremely important):
attr1_3
Attractive
sinc1_3
Sincere
intel1_3
Intelligent
fun1_3
Fun
amb1_3
Ambitious
shar1_3
Has shared interests/hobbies

Now, think back to your yes/no decisions during the night of the Speed Dating event. Try to distribute the 100 points among these six attributes in the way that best reflects the actual importance of these attributes in your decisions. Give more points to those attributes that were more important in your decisions, and fewer points to those attributes that less less important in your decisions. Total points must equal 100.
attr7_3
Attractive
sinc7_3
Sincere
intel7_3
Intelligent
fun7_3
Fun
amb7_3

Ambitious
shar7_3
Has shared interests/hobbies

Now we want to know what you think MOST of your fellow men/women look for in the opposite sex.
Please rate the importance of the following attributes on a scale of 1-10 (1=not at all important, 10=extremely important):
attr4_3
Attractive
sinc4_3
Sincere
intel4_3
Intelligent
fun4_3
Fun
amb4_3
Ambitious
shar4_3
Has shared interests/hobbies

What do you think the opposite sex looks for in a date?
Please rate the importance of the following attributes on a scale of 1-10 (1=not at all important, 10=extremely important):
attr2_3
Attractive
sinc2_3
Sincere
intel2_3
Intelligent
fun2_3
Fun
amb2_3
Ambitious
share2_3
Has shared interests/hobbies

Please rate your opinion of your own attributes, on a scale of 1-10 (1= awful and 10=great). Be honest!
attr3_3
Attractive
sinc3_3
Sincere
intel3_3

Intelligent
fun3_3
Fun
amb3_3
Ambitious

And finally, how do you think others perceive you?
Please rate yourself how you think others would rate you on each of the following attributes, on a scale of 1-10 (1=awful, 10=great)
attr5_3
Attractive
sinc5_3
Sincere
int5_3
Intelligent
fun5_3
Fun
amb5_3
Ambitious