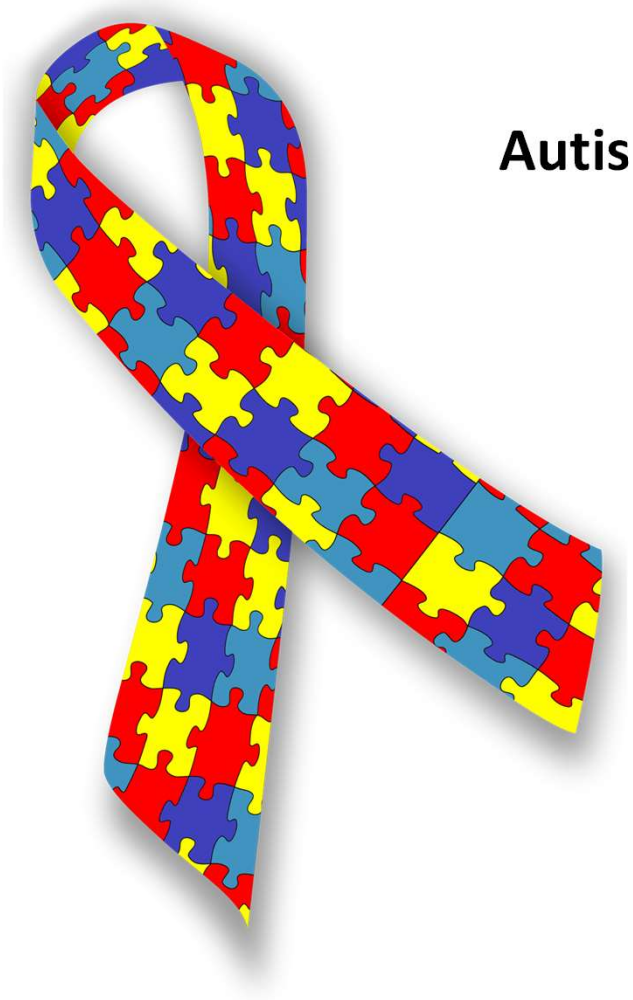


Autism Spectrum Disorder Prediction

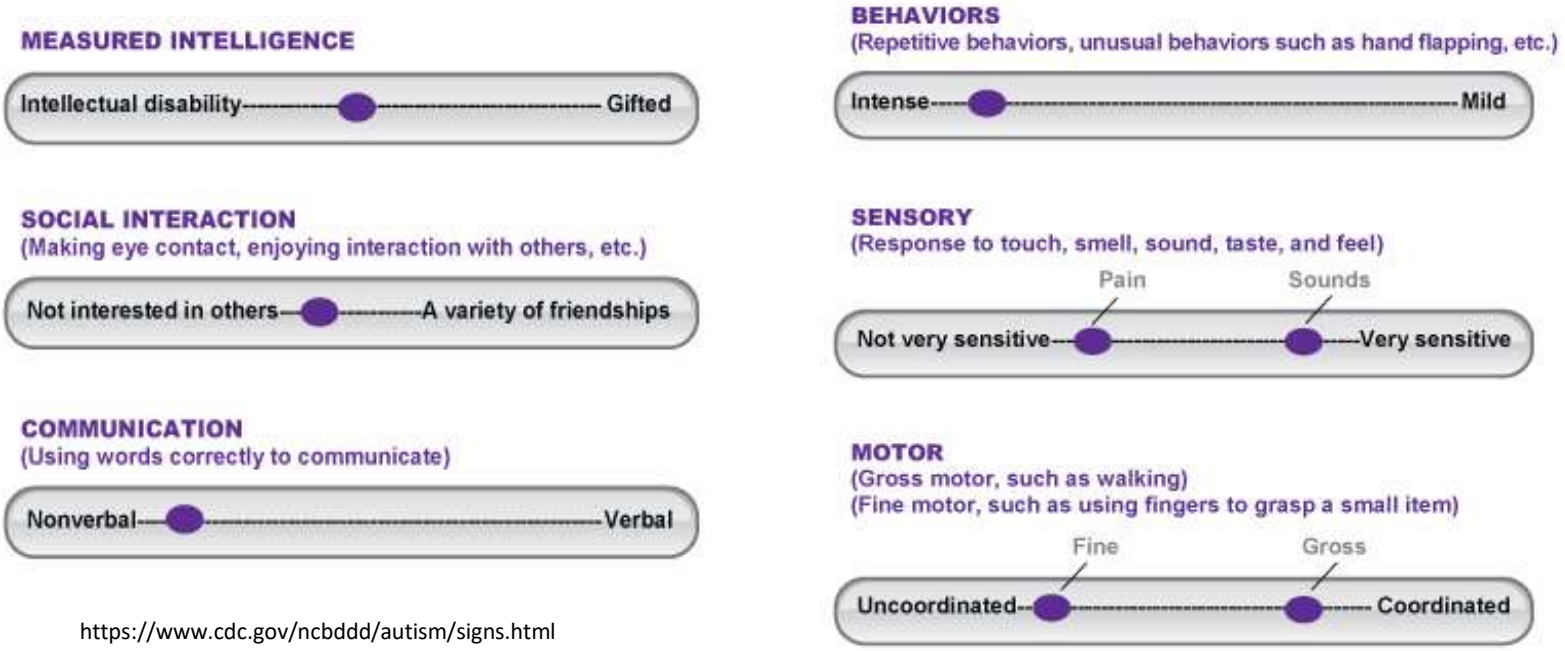
Yimiao Ou

Autism Spectrum Disorder (ASD or autism):

- A neurodevelopmental disorder that impacts brain development.
- Affecting individual's learning ability, interaction and communication with others.
- It begins early in childhood and lasts throughout an individual's life.
- About 1 in 68 children has been identified with ASD.

Centers for Disease Control and Prevention (CDC) https://www.cdc.gov/ncbddd/autism/index.html

Range of Symptoms



https://www.cdc.gov/ncbddd/autism/signs.html

Diagnosis 7
Co-occurring / Comorbid Conditions
Evolution of Autism

The Adult Autism Spectrum Quotient (AQ)

AGES 16 +

How to fill out the questionnaire

Below are a list of statements. Please read each statement very carefully and rate how strongly you agree or disagree.

		Definitely Agree	Slightly Agree	Slightly Disagree	Definitely Disagree
1	I prefer to do things with others rather than on my own.	○	○	○	○
2	I prefer to do things the same way over and over again.	○	○	○	○
3	If I try to imagine something, I find it very easy to create a picture in my mind.	○	○	○	○
4	I frequently get so strongly absorbed in one thing that I lose sight of other things.	○	○	○	○
5	I often notice small sounds when others do not.	○	○	○	○
6	I usually notice car number plates or similar strings of information.	○	○	○	○
7	Other people frequently tell me that what I've said is impolite, even though I think it is polite.	○	○	○	○
8	When I'm reading a story, I can easily imagine what the characters might look like.	○	○	○	○

BBC Sign in | News | Sport | Weather | Capital | TV | Radio | More... | Search BBC News

NEWS

Watch ONE-MINUTE WORLD NEWS

News Front Page

Africa
Americas
Asia-Pacific
Europe
Middle East
South Asia
UK
Business
Health
Medical notes
Science & Environment
Technology
Entertainment
Also in the news

Video and Audio

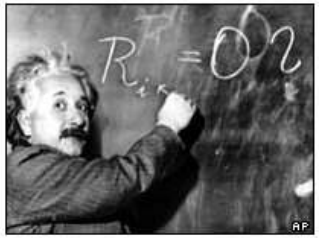
Page last updated at 23:13 GMT, Wednesday, 30 April 2003 00:13 UK

E-mail this to a friend Printable version

Einstein and Newton 'had autism'

Albert Einstein and Isaac Newton may have suffered from a type of autism, according to experts.

Researchers at Cambridge and Oxford universities believe both scientists displayed signs of Asperger's Syndrome.



Einstein was a notoriously confusing lecturer

Many people with Asperger's are often regarded as being eccentric. They sometimes lack social skills, are obsessed with complex topics and can have problems communicating.

This latest research suggests that Einstein, who is credited with developing the theory of relativity, and Newton, who discovered the laws of gravity, had these traits to varying degrees.

SEE ALSO
- Time for Einstein
  26 Dec 99 | Science/Nature
- Why size mattered for Einstein
  18 Jun 99 | Science/Nature
- Asperger syndrome
  25 Aug 09 | Health

RELATED BBC LINKS
- Isaac Newton biography

RELATED INTERNET LINKS
- New Scientist
- University of Cambridge
- University of Oxford

The BBC is not responsible for the content of external internet sites

Can we correctly predict autism based on signs and symptoms?

What are the most important symptoms(features) in autism prediction?

Dataset: Autism Screening Adult Data Set
https://archive.ics.uci.edu/ml/datasets/Autism+Screening+Adult

Data mining software: Scikit-Learn(python), RapidMiner

Data mining functions:

Classification: (autism/not autism)
- Decision Tree
- Naïve Bayes
- Logistic Regression
- Support Vector Machine
- Feed-forward Neural Network
- k-NN

Association rules(RapidMiner)

Data attributes

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	age	gender	ethnicity	jaundice	autism	country_of_res	used_app_before	result	age_desc	relation	class
1	1	1	1	0	0	1	1	0	0	26	f	White-European	no	no	'United States'	no	6	'18 and more'	Self	NO
1	1	0	1	1	0	1	1	1	1	27	m	Latino	yes	yes	Spain	no	8	'18 and more'	Parent	YES

A1: I often notice small sounds when others do not.
A2: I usually concentrate more on the whole picture, rather than the small details.
A3: I find it easy to do more than one thing at once.
A4: If there is an interruption, I can switch back to what I was doing very quickly.
A5: I find it easy to read between the lines when someone is talking to me.
A6: I know how to tell if someone listening to me is getting bored.
A7: When I am reading a story I find it difficult to work out the character's intentions.
A8: I like to collect information about categories of things.
A9: I find it easy to work out what someone is thinking or feeling just by looking at their face.
A10: I find it difficult to work out people's intentions.
age: age in years
gender: male(m) or female(f)
ethnicity: string
jaundice: whether the case was born with jaundice(yes/no)
autism: whether any immediate family member has autism(yes/no)
country_of_res: country of residence
used_app_before: used the screening app before(yes/no)
result: final score(total number of 1) obtained from A1 to A10
age_desc: string
relation: who is completing the test
class: target label(YES/NO)

Dataset:
704 samples

Data preprocessing

data = pandas.read_csv(file_path)

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	age	gender	ethnicity	jaundice	autism	country_of_res	used_app_before	result	age_desc	relation	class
1	1	1	1	0	0	1	1	0	0	26	f	White-European	no	no	'United States'	no	6	18 and more'	Self	NO
1	1	0	1	1	0	1	1	1	1	27	m	Latino	yes	yes	Spain	no	8	'18 and more'	Parent	YES
						...	...	...	...	...										
1	0	0	0	0	0	0	0	0	0	383	f	Pasifika	no	no	'New Zealand'	no	1	'18 and more'	Self	NO

Remove samples with missing value — data.replace('?', numpy.nan); data.dropna(axis = 0, inplace = True)

Remove non-informative attributes — data.drop(['ethnicity'], 1)

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	age	gender	jaundice	autism	class
1	1	1	1	0	0	1	1	0	0	26	f	no	no	NO
1	1	0	1	1	0	1	1	1	1	27	m	yes	yes	YES
						...	...	...	...	...				
1	0	0	0	0	0	0	0	0	0	383	f	no	no	NO

Replace abnormal value with average value — data.loc[data.age == 383, 'age'] = 30

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	age	gender	jaundice	autism	class
1	1	1	1	0	0	1	1	0	0	26	f	no	no	NO
1	1	0	1	1	0	1	1	1	1	27	m	yes	yes	YES
						...	...	...	...	...				
1	0	0	0	0	0	0	0	0	0	30	f	no	no	NO

Change attributes to numeric — data.replace({'gender' : 'f'}, 1).replace({'gender' : 'm'}, 0)

A1	A2	A3	A4	A5	A6	A7	A8	A9	A10	age	gender	jaundice	autism	class
1	1	1	1	0	0	1	1	0	0	26	1	0	0	0
1	1	0	1	1	0	1	1	1	1	27	0	1	1	1
						...	...	...	...	...				
1	0	0	0	0	0	0	0	0	0	30	1	0	0	0

Scikit-Learn

Read csv file

```
data = pandas.read_csv(file_path)
dataX = data.drop(['class'], 1)
dataY = data['class']
```

Split data into training set and test set

```
x_train, x_test, y_train, y_test = train_test_split(dataX, dataY, test_size = 0.3, random_state = 40)
```

Normalize data(optional)

```
scaler = StandardScaler()
scaler.fit(x_train)
x_train = scaler.transform(x_train)
x_test = scaler.transform(x_test)
```

Parameter optimization with 10-fold cross validation

```
svm = SVC(C=1, kernel = 'linear', degree = 1, cache_size = 200, max_iter = 100000)
scores = cross_val_score(svm, x_train, y_train, cv = 10)
print scores.mean(), scores.std()*2
```

Model fitting and test data prediction

```
svm.fit(x_train, y_train)
prediction = svm.predict(x_test)
accuracy = svm.score(x_test, y_test)
```

Classification Result

Scikit-Learn	Decision Tree	Naïve Bayes	Logistic Regression	Support Vector Machine	Neural Network	k-NN
Accuracy	0.92	0.94	1	1	0.99	0.98
Precision	0.94	0.89	1	1	0.98	0.98
Recall	0.81	0.93	1	1	1	0.95
F1-score	0.87	0.91	1	1	0.99	0.97

RapidMiner	Decision Tree	Naïve Bayes	Logistic Regression	Support Vector Machine	Neural Network	k-NN
Accuracy	0.91	0.95	1	1	1	0.95
Precision	0.88	0.92	1	1	1	0.89
Recall	0.81	0.91	1	1	1	0.94
F1-score	0.84	0.92	1	1	1	0.91

RapidMiner

Decision Tree

```
A9 = 0
|   A6 = 0: 0 {0=225, 1=8}
|   A6 = 1
|   |   A5 = 0: 0 {0=14, 1=1}
|   |   A5 = 1
|   |   |   age > 19.500
|   |   |   |   A2 = 0
|   |   |   |   |   A7 = 0: 0 {0=7, 1=0}
|   |   |   |   |   A7 = 1
|   |   |   |   |   |   A4 = 0: 0 {0=2, 1=0}
|   |   |   |   |   |   A4 = 1: 1 {0=0, 1=3}
|   |   |   |   A2 = 1
|   |   |   |   |   A1 = 0
|   |   |   |   |   |   gender = 0: 0 {0=2, 1=0}
|   |   |   |   |   |   gender = 1: 1 {0=0, 1=2}
|   |   |   |   |   A1 = 1: 1 {0=0, 1=10}
|   |   |   age ≤ 19.500: 0 {0=3, 1=0}
A9 = 1
|   A5 = 0
|   |   A3 = 0: 0 {0=18, 1=0}
|   |   A3 = 1
|   |   |   A7 = 0
|   |   |   |   age > 35: 1 {0=1, 1=2}
|   |   |   |   age ≤ 35: 0 {0=9, 1=0}
|   |   |   A7 = 1: 1 {0=0, 1=3}
|   A5 = 1
|   |   A6 = 0
|   |   |   age > 19.500
|   |   |   |   A8 = 0: 0 {0=6, 1=1}
|   |   |   |   A8 = 1
|   |   |   |   |   A4 = 0
|   |   |   |   |   |   A3 = 0: 0 {0=4, 1=0}
|   |   |   |   |   |   A3 = 1
|   |   |   |   |   |   |   A7 = 0: 0 {0=1, 1=1}
|   |   |   |   |   |   |   A7 = 1: 1 {0=0, 1=2}
|   |   |   |   |   A4 = 1
|   |   |   |   |   |   A1 = 0: 0 {0=2, 1=1}
|   |   |   |   |   |   A1 = 1: 1 {0=0, 1=16}
|   |   |   age ≤ 19.500: 0 {0=2, 1=0}
|   |   A6 = 1
|   |   |   A10 = 0
|   |   |   |   A8 = 0: 0 {0=3, 1=1}
|   |   |   |   A8 = 1
|   |   |   |   |   jundice = 0: 1 {0=0, 1=7}
|   |   |   |   |   jundice = 1: 0 {0=1, 1=1}
|   |   |   A10 = 1: 1 {0=0, 1=67}
```

A5, A6, A9, A10

Association Rules

[A5, class] --> [A9] (confidence: 0.807)
[class] --> [A9] (confidence: 0.811)
[A6] --> [A10] (confidence: 0.818)
[A6] --> [A5] (confidence: 0.824)
[A10, A9] --> [A5] (confidence: 0.831)
[A5, A9] --> [A10] (confidence: 0.831)
[A5, A9] --> [class] (confidence: 0.831)
[class] --> [A10, A5] (confidence: 0.844)
[class] --> [A10] (confidence: 0.883)
[A5, class] --> [A10] (confidence: 0.889)
[A9, class] --> [A5] (confidence: 0.945)
[class] --> [A5] (confidence: 0.950)
[A10, class] --> [A5] (confidence: 0.956)

Classification result with attributes A5, A6, A9, A10

Scikit-Learn	Decision Tree	Naïve Bayes	Logistic Regression	Support Vector Machine	Neural Network	k-NN
Accuracy	0.91	0.81	0.9	0.9	0.9	0.87
Precision	0.91	0.83	0.91	0.91	0.91	0.88
Recall	0.91	0.81	0.9	0.9	0.9	0.87
F1-score	0.91	0.81	0.9	0.9	0.9	0.87

Attention to detail
- A1: I often notice small sounds when others do not.
- A2: I usually concentrate more on the whole picture, rather than the small details.

Attention switching
- A3: I find it easy to do more than one thing at once.
- A4: If there is an interruption, I can switch back to what I was doing very quickly.

Communication
- A5: I find it easy to read between the lines when someone is talking to me.
- A6: I know how to tell if someone listening to me is getting bored.

Imagination
- A7: When I am reading a story I find it difficult to work out the character's intentions.
- A8: I like to collect information about categories of things.

Social interaction
- A9: I find it easy to work out what someone is thinking or feeling just by looking at their face.
- A10: I find it difficult to work out people's intentions.

Conclusions

With attributes A1, A2, A3, A4, A5, A6, A7, A8, A9, A10, age, gender, jaundice and relative_has_autism, we can predict autism with high accuracy using Decision Tree, Naïve Bayes, Logistic Regression, Support Vector Machine, Feed-forward Neural Network and k-NN.

Models based on A5, A6, A9 and A10 alone produced comparable accuracy to models using more attributes, suggesting that communication and social interaction are two important factors for correctly predicting autism.

Dataset: Autism Screening Adult Data Set

Dataset: Autistic Spectrum Disorder Screening Data for Children Data Set

Accuracy (with attributes A1, A2, A3, A4, A5, A6, A7, A8, A9, A10, age)

	Decision Tree	Naïve Bayes	Logistic Regression	Support Vector Machine	Neural Network	k-NN
Scikit-Learn	0.9	0.76	1	1	1	0.9
RapidMiner	0.88	0.94	1	1	1	0.88

Accuracy (with attributes A3, A4, A7, A9, A10)

	Decision Tree	Naïve Bayes	Logistic Regression	Support Vector Machine	Neural Network	k-NN
Scikit-Learn	0.88	0.68	0.88	0.88	0.88	0.88

Attention to detail
- A1: S/he often notices small sounds when others do not.
- A2: S/he usually concentrates more on the whole picture, rather than the small details.

Attention switching
- A3: In a social group, s/he can easily keep track of several different people's conversations.
- A4: S/he finds it easy to go back and forth between different activities.

Communication
- A5: S/he doesn't know how to keep a conversation going with his/her peers.
- A6: S/he is good at social chit-chat.

Imagination
- A7: When s/he is reading a story, s/he finds it difficult to work out the character's intentions or feelings.
- A8: When s/he was in preschool, s/he used to enjoy playing games involving pretending with other children.

Social interaction
- A9: S/he finds it easy to work out what someone is thinking or feeling just by looking at their face.
- A10: S/he finds it hard to make new friends.

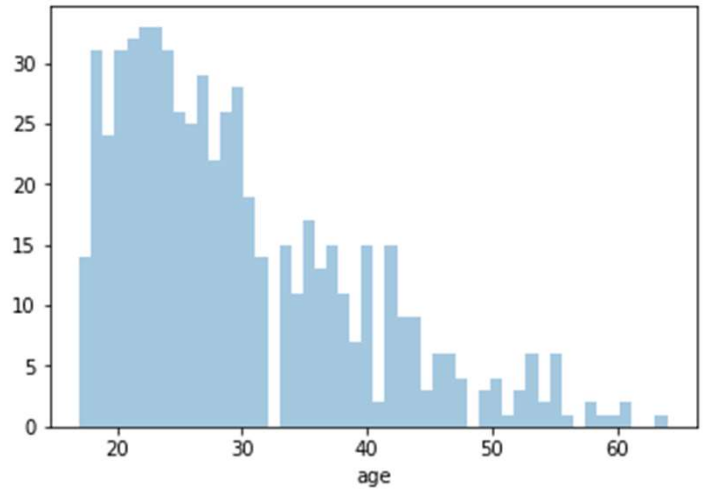
Future Work and Challenges

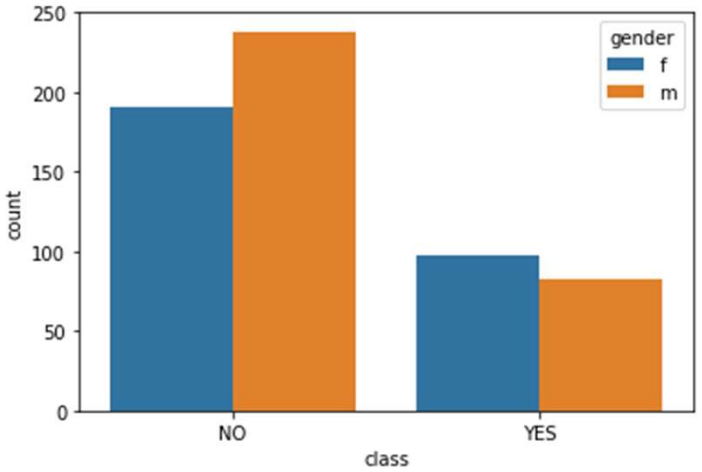
1. How well do the models generalize to data from other sources?

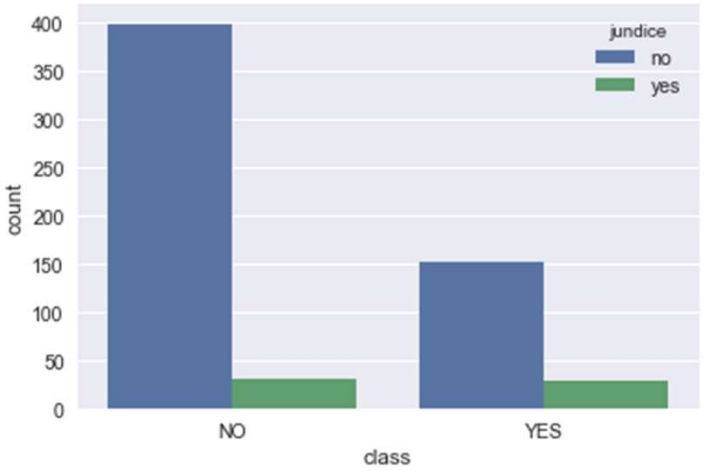
2. Where to get a larger data set? Finding a larger **clinical** data set would be difficult or might not be possible.

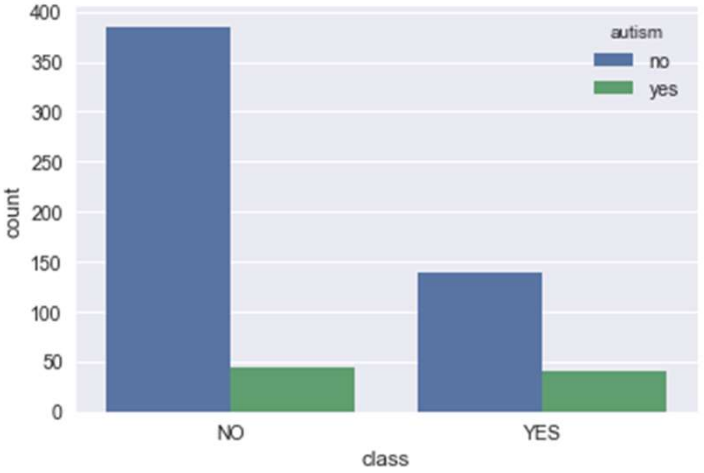
3. Data preprocessing was another challenge. How to handle missing data?

Thank you!



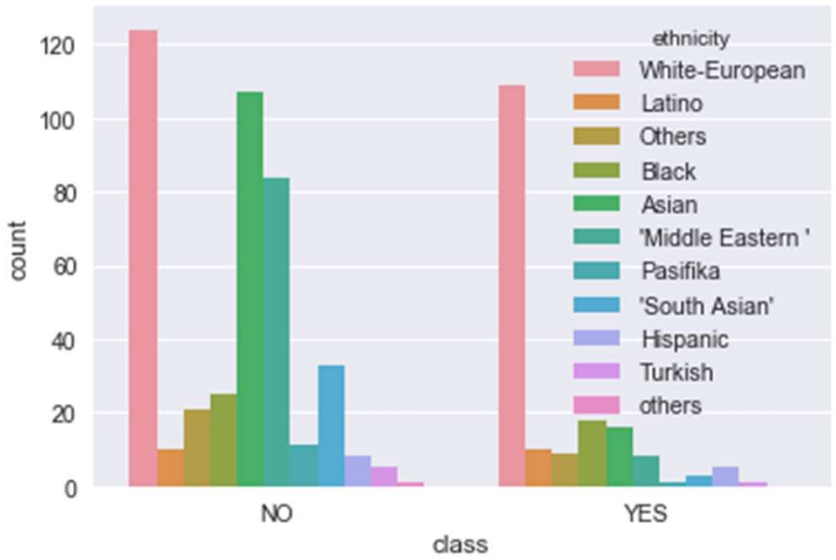


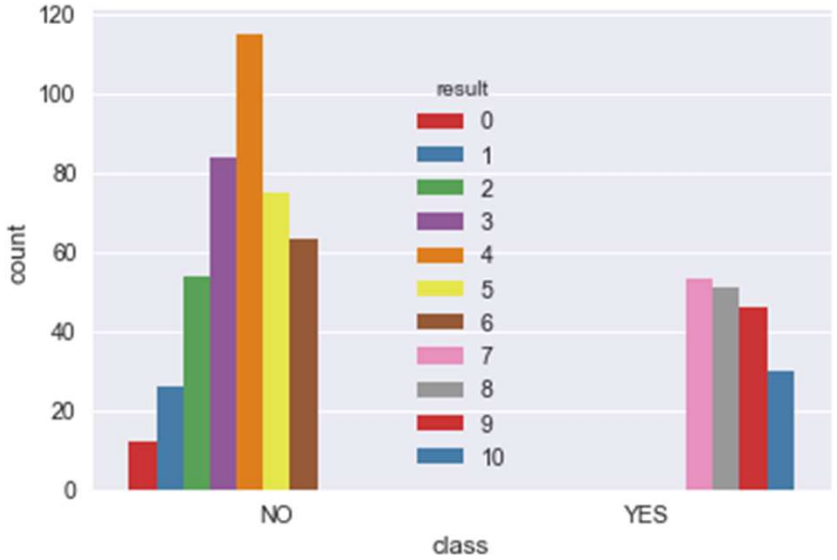




seaborn. distplot(data['age'], bins = 50, kde = False)

seaborn.countplot(x = 'class', hue = ‘gender', data)


ethnicity
White-European
Latino
Others
Black
Asian
'Middle Eastern '
Pasifika
'South Asian'
Hispanic
Turkish
others
count
class
NO
YES
0
20
40
60
80
100
120


result
0
1
2
3
4
5
6
7
8
9
10
count
class
NO
YES
0
20
40
60
80
100
120

Scikit-Learn

Decision Tree

```
dcTree = tree.DecisionTreeClassifier()
```

Naïve Bayes

```
NB = MultinomialNB(alpha = 1, class_prior = None, fit_prior = True)
```

Logistic Regression

```
LR = LogisticRegression(penalty = 'l1')
```

Support Vector Machine

```
svm = SVC(C=1, kernel = 'linear', degree = 1, cache_size = 200, max_iter = 100000)
```

Feed_forward Neural Network

```
NN = MLPClassifier(hidden_layer_sizes = (10), max_iter = 1000, learning_rate_init = 0.05, momentum = 0.1)
```

K-NN

```
knn = KNeighborsClassifier(n_neighbors = 10)
```

RapidMiner with attributes A5, A6, A9, A10

Decision Tree

PerformanceVector:
accuracy: 84.70%
ConfusionMatrix:

True:	0	1
0:	113	12
1:	16	42

Logistic Regression

PerformanceVector:
accuracy: 85.79%
ConfusionMatrix:

True:	0	1
0:	115	12
1:	14	42

Support Vector Machine

PerformanceVector:
accuracy: 85.25%
ConfusionMatrix:

True:	0	1
0:	113	11
1:	16	43

```
A9 = 0
|   A6 = 0: 0 {0=225, 1=8}
|   A6 = 1
|   |   A5 = 0: 0 {0=14, 1=1}
|   |   A5 = 1: 1 {0=14, 1=15}
A9 = 1
|   A5 = 0: 0 {0=28, 1=5}
|   A5 = 1
|   |   A6 = 0
|   |   |   A10 = 0: 0 {0=6, 1=2}
|   |   |   A10 = 1: 1 {0=9, 1=19}
|   |   A6 = 1: 1 {0=4, 1=76}
```