

Handwriting Personality Prediction

Group B1: Kaushik Kumaran, Hyeon Gu (Grant) Kim, Lucy Hwang, Yashaswini Kalva, Archit Patel



Introduction

Overview

Applications

Psychological analysis

Graphology is used clinically by counsellors and psychotherapists.

Employment profiling

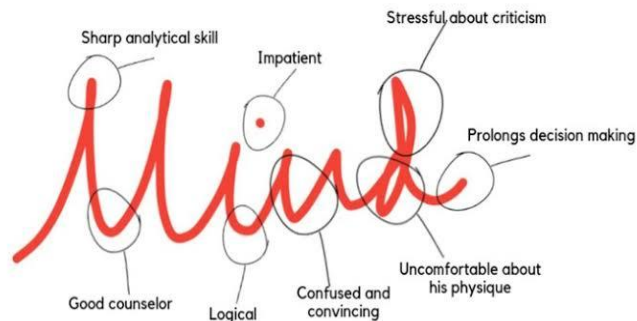
A graphological report is meant to be used in conjunction with other tools, such as comprehensive background checks, practical demonstration or record of work skills.

Problem Statement

A system is proposed to **automate** the basic handwriting analysis tasks of graphology to determine a few important **personality traits**.

Analysis example

Analysis example



Sample walkthrough

Sample text

Dr Sir,
Your note has just been sent to me
from Florence.
Here, therefore, is an Autograph
Yours Truly
Atkinson



1

Dr Sir,
Your note has just been sent to me
from Florence.
Here, therefore, is an Autograph
Yours Truly
Atkinson

Alignment and size of letter

2

Dr Sir,
Your note has just been sent to me
from Florence.
Here, therefore, is an Autograph
Yours Truly
Atkinson

Match the template

3

Dr Sir,
Your note has just been sent to me
from Florence.
Here, therefore, is an Autograph
Yours Truly
Atkinson

Spacing

Feature examples

Alignment and size

Mr Sir,
Your note has just been sent to me
from Florence.

Yours truly,

Alfred Russel
Wallace



Alignment

A quick brown fox

Ascending baseline

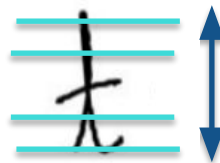
A quick brown fox

Descending baseline

A quick brown fox

Level baseline

Size



Data sourcing and cleaning

Data sourcing

Data from the **IAM Handwriting Database of Research Group** on Computer Vision and Artificial Intelligence INF, University of Bern, Switzerland is obtained.

IAM Handwriting Database of Research Group

672 pages of scanned text

Data cleaning



Data input



Preprocessing



Feature extraction



Training the classifier



Predict personality traits

Methods

Image Resolution and Cropping

The original images contain **unwanted printed texts, lines** and **free space** which are not suitable for further processing. We will be writing a script to **crop out** the left and right margins, **resize** all the images with 850 pixels width and perspective height, and save the images in PNG format.

Noise Removal

Image noise is defined as **random variation of brightness** or **colour** information in images. Noise can be removed by **filtering** the images.

Handwriting features

Attribute	Writing categories	Psychological personality behavior
Letter size	Large letters	Likes being noticed, stands out in a crowd
	Small letters	Introspective, not seeking attention, modest
	Average letters	Adaptable, fits into a crowd, practical, balanced
Letter slant	Right slant	Sociable, responsive, interested in others, friendly
	Left slant	Reserved, observant, self-reliant, non-intrusive
	Vertical slant	Practical, independent, controlled, self-sufficient
Pen pressure	Light pen pressure	Can endure traumatic experiences w/o seriously affected.
	Heavy pen pressure	Have very deep and enduring feelings
Baseline	Raising baseline	Optimistic, upbeat, positive attitude, ambitious and hopeful
	Falling baseline	Tired, overwhelmed, pessimistic, not hopeful
	Straight baseline	Determined, stays on track, self-motivated, reliable, steady
	Erratic baseline	Wavering, lacks definite direction, unpredictable
Word spacing	Far spaced words	Desires more space, enjoys privacy
	Close spaced words	Closeness of sentiment and intelligence

Labels

Agreeableness

Neuroticism

Openness

Consciousness

Extraversion



Modeling

Model overview

CNN

The input image is fed into the CNN layers. These layers are trained to extract relevant features from the image.

ANN

Computational models inspired by an animal's central nervous systems. It is capable of machine learning as well as pattern recognition

Random forest

An ensemble learning method for classification, regression and other tasks that operates by constructing a multitude of decision trees at training time.

CNN Methodology

Data Preprocessing

- Separated the data into training set, validation set, and test set samples
- **Data Augmentation** was used to increase the number of samples

Model Building

- **Transfer learning** was used to learn the lower level features
- The base model was **InceptionResnetv2** with pre-trained weights flowing in from **ImageNet** dataset
- The base model was followed by **Max Pooling, Dropout, Batch Normalization** and a fully connected layer of **50 units**

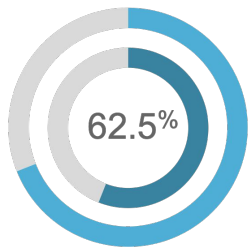
Model Features

- **Relu** activation function for the hidden layers
- **SoftMax** activation is used for the output layer
- **Adam Optimization** is used for Gradient Descent
- Incorporated **model checkpoints** to store the best weights

Hyperparameter Tuning

- **Epochs:** set the number of epochs to 30
- **Batch size:** tried batch sizes of 16, 32, and 64
- **Learning rate:** 0.001

CNN Results and Next Steps



Results

Accuracy on the training set - 78.9%
Accuracy on the validation set - 65.3%
Accuracy on the test set - 62.5%

Next steps

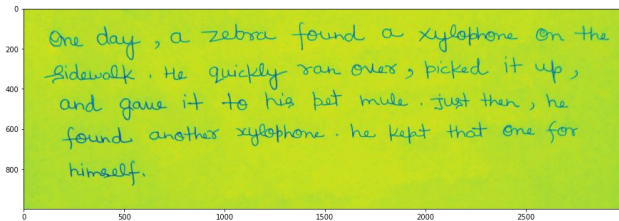
Unfreeze certain layers and try re training the model with our dataset for those layers

Try other architectures which could potentially outperform Inception Resnet V2 for the given dataset

Augment the data further for the imbalanced classes.
Also tune parameters like the optimizer, number of layers etc

ANN

Sample text

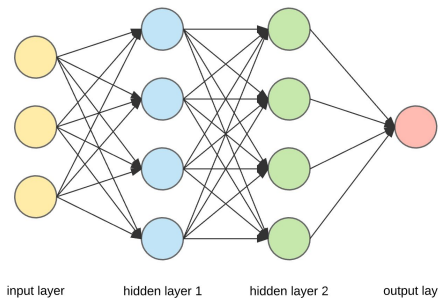


Arrays of Pixels

```
array([[174, 174, 174, ..., 173, 172, 173],  
       [173, 172, 173, ..., 172, 171, 172],  
       [174, 173, 173, ..., 173, 171, 172],  
       ...,  
       [162, 162, 163, ..., 166, 166, 165],  
       [162, 162, 162, ..., 166, 165, 165],  
       [162, 162, 162, ..., 165, 165, 164]], dtype=uint8)
```



ANN



ANN

Data Preprocessing

- Encoded the data
- Split the data into train, validation, and test data sets

Model Building

- Rescaled and flattened the data
- **ReLU activation function** for the hidden layers
- **Softmax** for the output layer
- Regularized each hidden layers to prevent overfitting.

Model Features

- **Sparse Categorical Cross entropy** loss function
- **RMSprop** optimizer

Hyperparameter Tuning

- **Epochs:** 60
- **Batch size:** tried batch sizes of 16, 32, and 64

ANN

Data Preprocessing

- Encoded the data
- Split the data into train, validation, and test data sets

Model Building

- Rescaled and flattened the data
- **ReLU activation function** for the hidden layers
- **Softmax** for the output layer
- Regularized each hidden layers to prevent overfitting.

Model Features

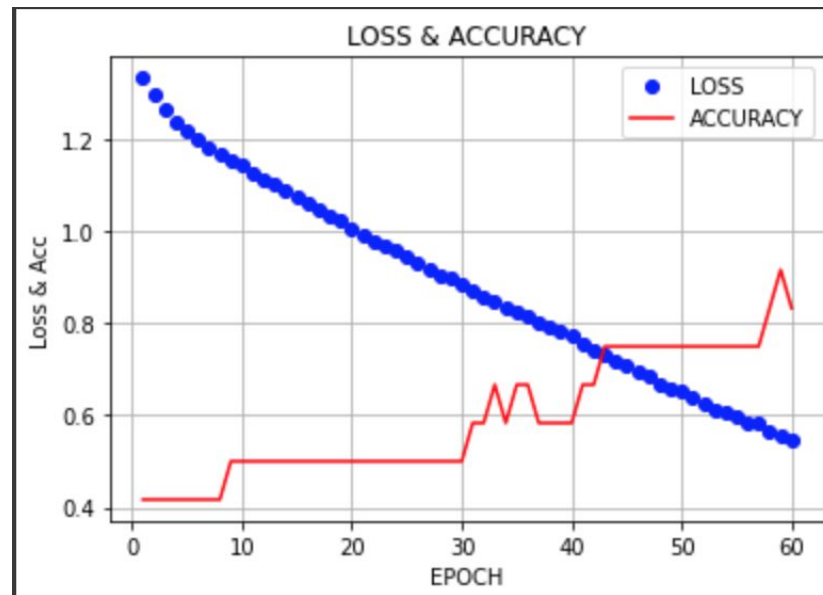
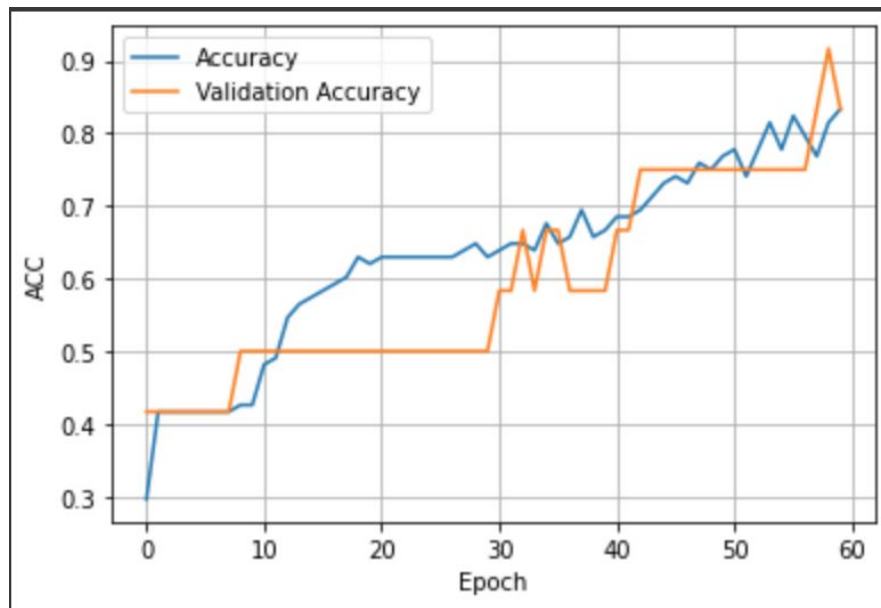
- **Sparse Categorical Cross entropy** loss function
- **RMSprop** optimizer

Hyperparameter Tuning

- **Epochs:** 60
- **Batch size:** tried batch sizes of 16, 32, and 64

Model accuracy

Accuracy on the training set - 83%
Accuracy on the validation set - 83.3%
Accuracy on the test set - 79%



ANN Challenges and Next Steps

Challenges

Tabular Data Input

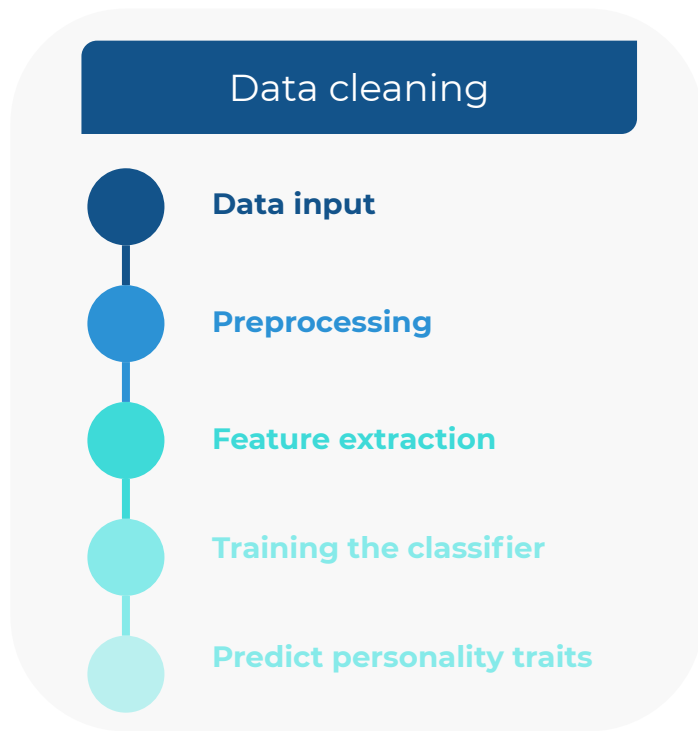
Hard to Interpret

Next Steps

Try Different
Hyperparameters

ADAM optimizer

Pre-processing and data cleaning



Pre-processing and data cleaning

The handwriting images we obtained contains unwanted noise, printed texts and lines

The aim of pre-processing is to make the image data suitable for feature extraction

No.	Pre-processing stages	Pre-processing steps
1	Image resolution and cropping	<ul style="list-style-type: none">• Crop left and right margins• Resize the images with 850 pixels width
2	Noise removal	<ul style="list-style-type: none">• Bilateral filter
3	Grayscale and binarization	<ul style="list-style-type: none">• Converted to grayscale• Binarized using inverted global thresholding
4	Contour and transformation	<ul style="list-style-type: none">• Straightening using dilation• Find the contours• Affine transformation
5	Horizontal and vertical projections	<ul style="list-style-type: none">• Horizontal projection: list of sum of all the pixel values of each row of the image• Vertical projection: list of sum of all the pixel values in each column of the image

Pre-Processing

Inverted Binarization

Mr Roy's United Federal Party is boycotting
the London talks on the Protectorate's
future, said Mr. Nkumbula last night;
"We want to discuss what to do if the

Dilation



Handwriting features

Attribute	Writing categories	Psychological personality behavior
Letter size	Large letters	Likes being noticed, stands out in a crowd
	Small letters	Introspective, not seeking attention, modest
	Average letters	Adaptable, fits into a crowd, practical, balanced
Letter slant	Right slant	Sociable, responsive, interested in others, friendly
	Left slant	Reserved, observant, self-reliant, non-intrusive
	Vertical slant	Practical, independent, controlled, self-sufficient
Pen pressure	Light pen pressure	Can endure traumatic experiences w/o seriously affected.
	Heavy pen pressure	Have very deep and enduring feelings
Baseline	Raising baseline	Optimistic, upbeat, positive attitude, ambitious and hopeful
	Falling baseline	Tired, overwhelmed, pessimistic, not hopeful
	Straight baseline	Determined, stays on track, self-motivated, reliable, steady
	Erratic baseline	Wavering, lacks definite direction, unpredictable
Word spacing	Far spaced words	Desires more space, enjoys privacy
	Close spaced words	Closeness of sentiment and intelligence

Feature Extraction

Baseline

Letter Size

Line Spacing

Word Spacing

Top Margin

Pen Pressure

Slant of Letters

Baseline

Normal

At the beginning of this wonderful century many people believed that there were no more worlds to

Rising

All types of trader have been encouraged, from the large departmental store to the small shoe-mender, with banks specially

Falling

But in 1897, alas! there was no fingerprint bureau, no experts to check and photograph any 'dabs' it

Affine & Warp Transformation

Before

It seems obvious that if an expedition to Brittany was compelled to attack via Colai's, then the primary essential to the success of the French war was a unity in unquestioned command of the Channel. It seems obvious that divided forces were dissipating the advantage of a ring of bridge-heads which included Colai's, Clorbourg, Bicot, Bordenas and Bayonne, and that there was no hope of final victory without a large-scale and concentrated invasion.

After

It seems obvious that if an expedition to Brittany was compelled to attack via Colai's, then the primary essential to the success of the French war was a unity in unquestioned command of the Channel. It seems obvious that divided forces were dissipating the advantage of a ring of bridge-heads which included Colai's, Clorbourg, Bicot, Bordenas and Bayonne, and that there was no hope of final victory without a large-scale and concentrated invasion.

Extracting Lines

Horizontal Projection

Commit to the lord whatever you do
and your plans will succeed

Letter Size

Normal size

'Not at all. I'll show you.' Oh no,
young Lee protested, considerate as always.

Large size

He did well. He got in touch with the
woman Pete was passing off as his

Small size

The large attendance and atmosphere of
this conference, held in October, 1957,

Line Spacing

Normal Spacing

It was locked. Egge's fingers felt bulky corkscrews and when he shook the case there was a rustling thud

Crowded Spacing

We did well. He got in touch with the woman Pete was passing off as his

Large Spacing

5. Other forms of war damage payment made by the Commission are highway payments, clearance payments (for clearing remains of

Word Spacing

It seems obvious that if an expedition to

Normal Spacing

Lee, however, showed not the slightest sign of fatigue. 'I wonder if you'd mind if I took a bit of a look round,'

Narrow Spacing

In many cases, their homes know little of their place of work and their associates at

Wide Spacing

Prof. Datta, who in 1911 was presented with a book on the life of Marx,

Top Margin



Top Margin

Mr. Macleod went on with the conference at Lancaster House despite the crisis which had blown up. He has now revealed his full plans to the Africans and Liberals attending. These plans do not give the Africans the overall majority they are seeking. African delegates are studying them today. The conference will meet to discuss the function of a proposed House of Chiefs.

Pen Pressure

Normal

Though they may gather some Left-wing support, a large majority of Labour M P's are likely to turn down the Foot-Griffiths resolution. Mr Foot's line will be that as Labour M P's opposed the Government Bill which brought life peers into existence, they should not now put forward nominees. He believes that the House of Lords should be abolished and that Labour should not take any steps which would appear to 'prop up' an out-dated institution.

Inverted

Though they may gather some Left-wing support, a large majority of Labour M P's are likely to turn down the Foot-Griffiths resolution. Mr Foot's line will be that as Labour M P's opposed the Government Bill which brought life peers into existence, they should not now put forward nominees. He believes that the House of Lords should be abolished and that Labour should not take any steps which would appear to 'prop up' an out-dated institution.

Slant of Letters

Inclined

"The whole naval might of the Soviet Union can be seen here - a truly inspiring

Normal

A submariner's wife needed to be spared as much as possible. Anyway the Parsifal affair was far too fresh in both their minds to be a comfortable

Reclined

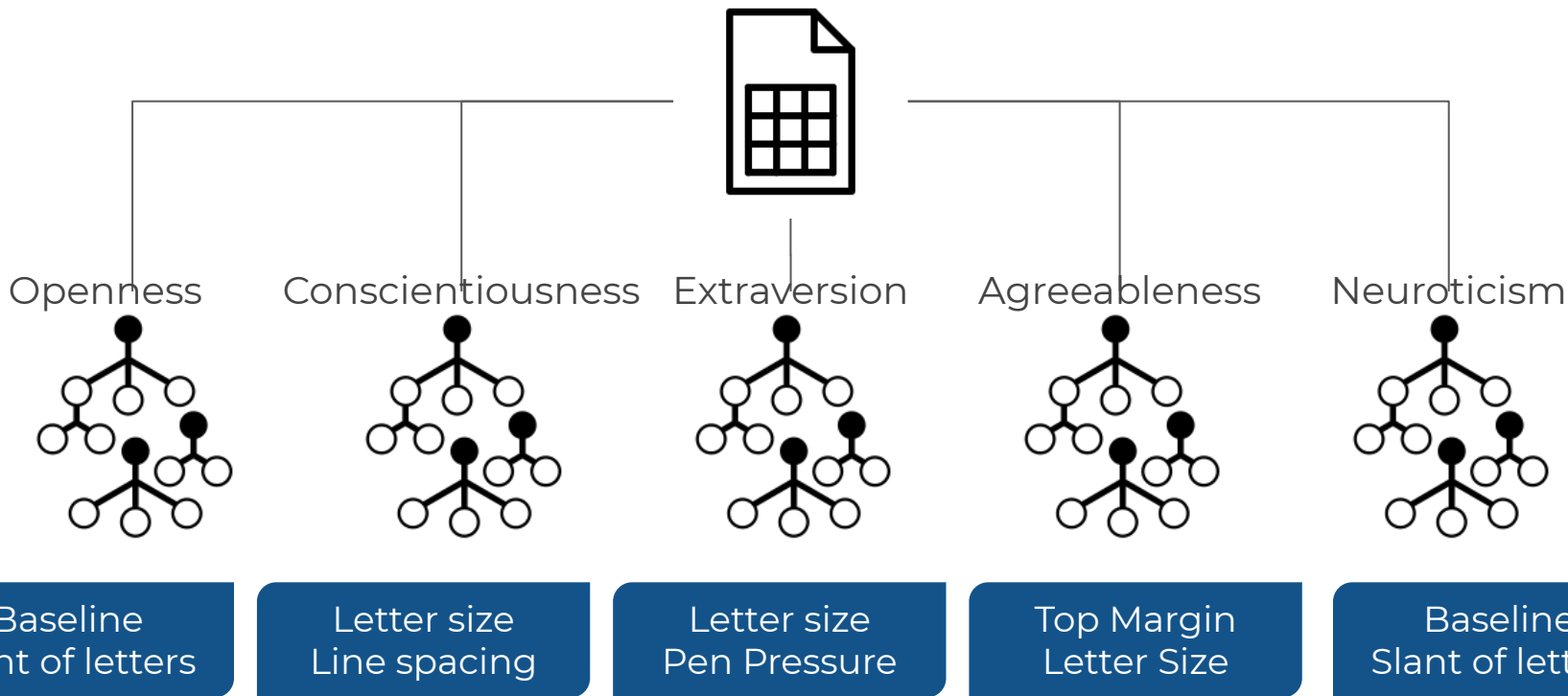
'I think it's gone beyond that, Prime Minister.

They feel pretty strongly about Africa. They are greatly disturbed by ~~the~~ the new and rather ugly image of

Processed Data

Baseline	Letter Size	Line Spacing	Word Spacing	Top Margin	Pen Pressure	Slant of Letters
-0.39	5.7	12.29	4.37	2.46	171.38	-15
0.05	3.07	17.29	2.89	1.74	194.71	-15
-1.1	2.02	16.36	3.14	1.65	170.29	-15
-0.01	1.91	15.73	3.32	1.6	165.56	45
0	2.1	15.7	3.25	1.69	171.55	-15
-0.27	2.14	16.33	3.16	1.54	174.16	-15
0.29	2.62	14.1	3.82	1.93	170.93	-15
-0.06	2.8	14.3	3.58	1.92	171.9	-15
0.11	2.43	14	3.78	1.78	171.28	45
-0.19	2.74	13.88	3.79	1.95	169.14	-15
0.13	2.8	14.63	3.94	1.93	167.94	-15
-0.39	3.08	14.6	3.66	1.92	169.14	-15
-0.16	1.96	17.36	2.94	1.38	182.99	-15
-0.26	2.22	15.33	3.41	1.76	169.25	30
-0.47	2.7	14.8	3.56	1.69	170.08	-15
-1.04	2.68	14.56	3.58	1.66	171.68	-15
-0.76	3.54	13	4.04	1.77	162.82	30
-0.33	3.04	14.82	3.62	1.91	158.97	-15
-0.23	2.63	14.82	3.58	1.74	168.45	-15
-0.02	3.29	14	3.81	1.72	167.25	-15
-0.7	2.73	15	3.59	1.76	160.74	45

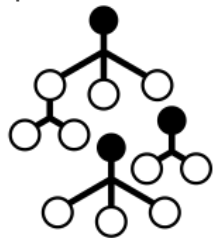
Random Forest



Random Forest

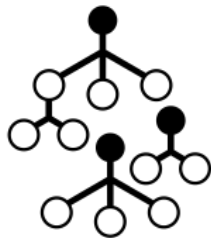


Openness



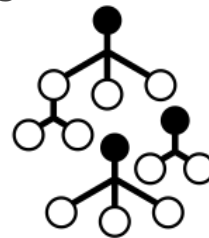
Baseline
Slant of letters

Conscientiousness



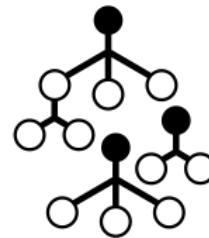
Letter size
Line spacing

Agreeableness



Top Margin
Letter Size

Neuroticism



Baseline
Slant of letters



Conclusion

Key takeaways

Pre-processing is a key step for accuracy

Hyper parameter tuning importance

Insights from our data

Key limitations

Extreme cases

Sample limitation

ANN trainable parameters

Looking forward

Next steps:

Neural Net to Extract Features

ROC

CNN next steps

Add more features

Plan to test professor's handwriting

Looking forward

To-do list

ROC

Computation and manual effort

Neural Net to Extract Features

Plan to test professor's handwriting

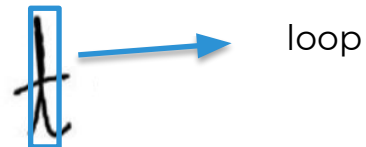
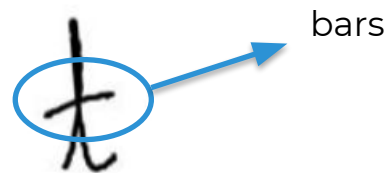
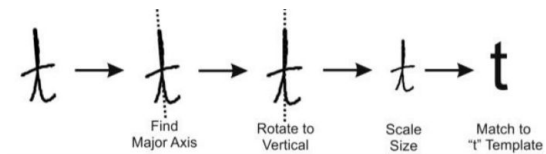
Thank you, Questions?



Appendix

Feature examples

Template matching



Overview

Applications

Psychological analysis

Graphology is used clinically by counsellors and psychotherapists.

Employment profiling

A graphological report is meant to be used in conjunction with other tools, such as comprehensive background checks, practical demonstration or record of work skills.

Problem Statement

A system is proposed to **automate** the basic handwriting analysis tasks of graphology to determine a few important **personality traits**.

Seven features/characteristics of a handwriting are considered to be extracted from a sample handwriting image.

Each of the seven resulting **raw values** will be put into **corresponding categories** of respective feature variations. A **combination** of these discrete values will be used to train each **support vector machine** for a **personality trait**. The classifier will then be able to predict the personality traits of the writer