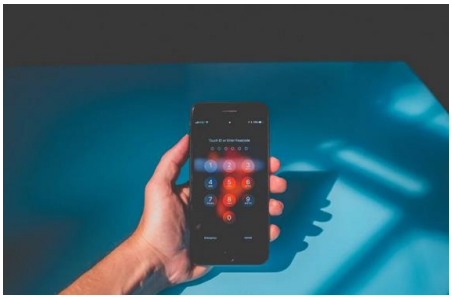
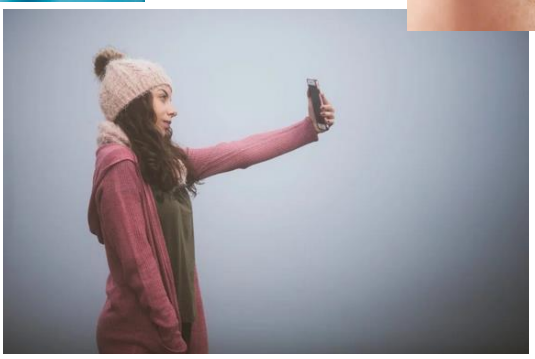
- Hi, I think that we have a good time to start the presentation, so let's do it
- From Internet Access Devices Usage to Behavioural Model – I will want to introduce you into this, in my humble opinion quite interesting, area of machine learning development

What is it?

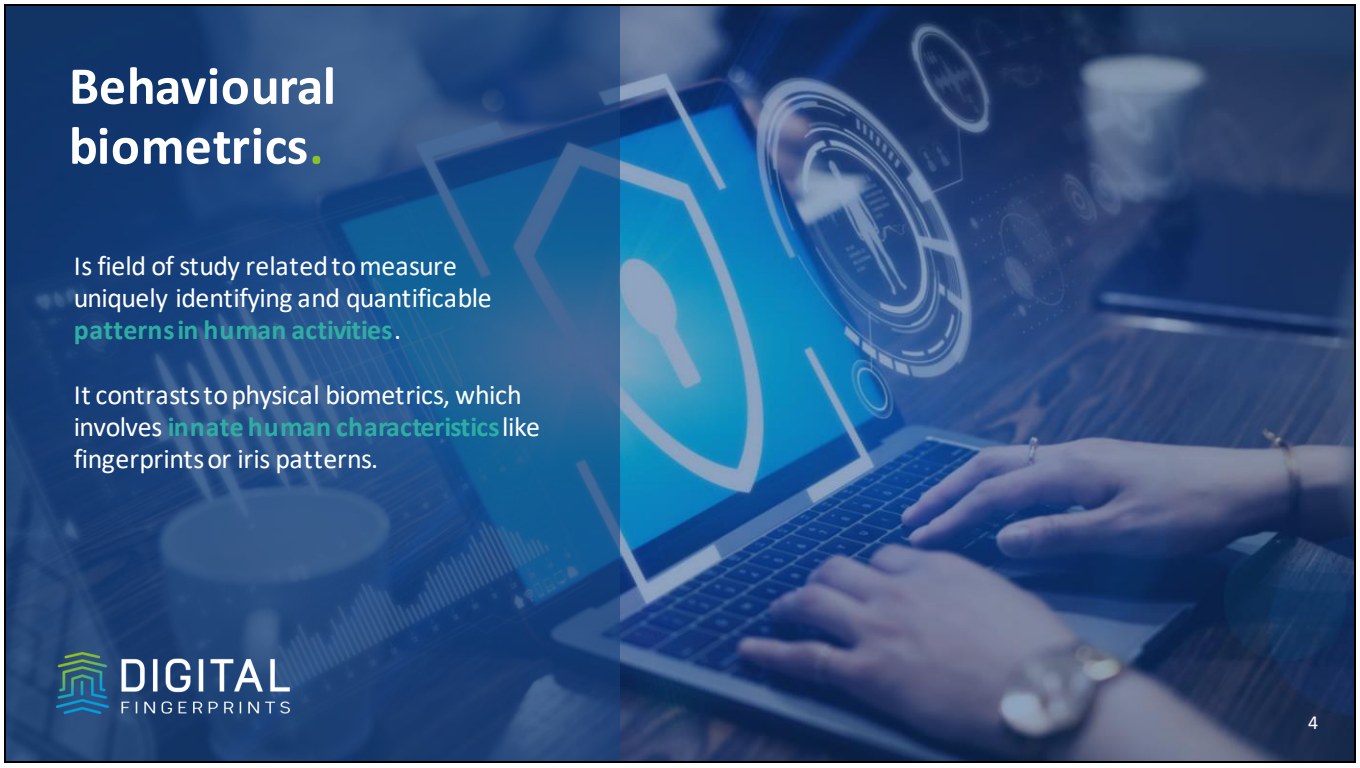
- Size: 20x14x8 mm.
- Weight: 5g.

- There is a puzzle – what is it?
- This thing is really small (by the way, we live in the era of miniaturization, so why it should be big?) and light (5 grams!)
- Any ideas?
-It is fingerprint reader working through USB. Cost – around 20 euros. So it is not an expensive thing, can be commonly used as **authentication tool**

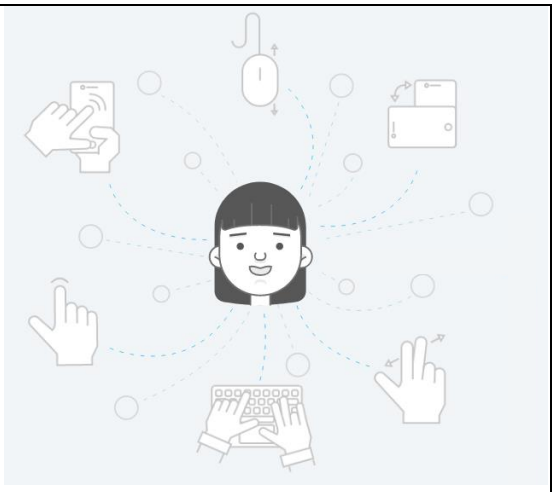




-Fingerprint is just first example of authentication method. Others are PIN or password, face recognition or iris recognition
-All of them require user engagement. User has to be aware what is doing. If he or she putting PIN or password or using facial/iris recognition.
-I don't want to conduct a business discussion about pros and cons of these methods. Just remember, that



-*read* - Is field of study related to measure uniquely identifying and measurable **patterns in human activities**.
-*read* - It contrasts to physical biometrics, which involves **innate human characteristics** like fingerprints or iris patterns.
-The main difference between behavioural and physical biometrics is that the first one does not require that user is aware of authentication process. User can make normal activities and the way in which he is doing these activities is used to conduct authentication.
- The main field in which behavioural biometrics is developing is the way of computer use with external devices like keyboard, mouse, touchpad.



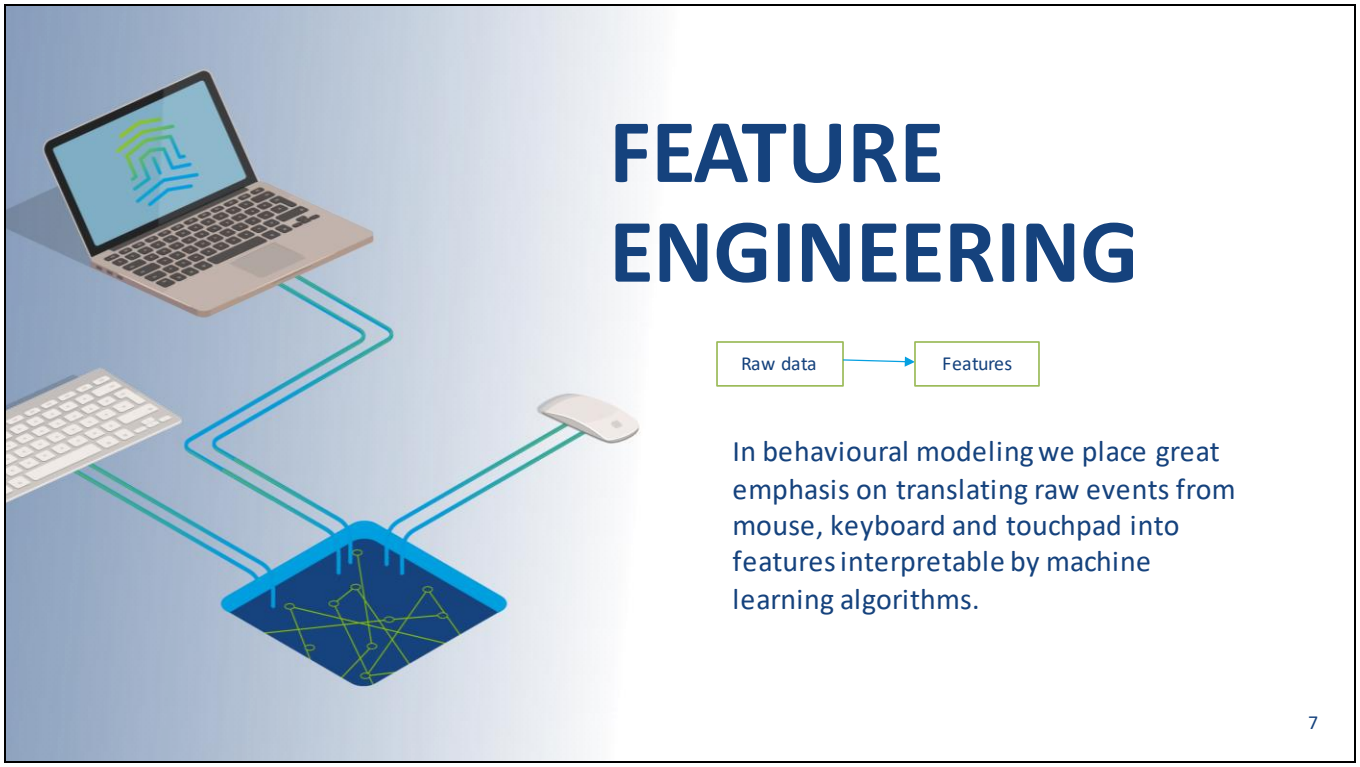
DATA COLLECTION

One man to rule them all.
All of them to authenticate one man.

-An average man uses many devices each day. Keyboard, mouse, tablet, mobile phone. When we make something regularly, we are making habits. So each of us has his own, specific way of using these devices.
-Just like with human fingerprint, our behaviour is unique. The way we pressing keys, moving mouse or swap the screen is characteristic for each of us. So it can be used as authentication tool

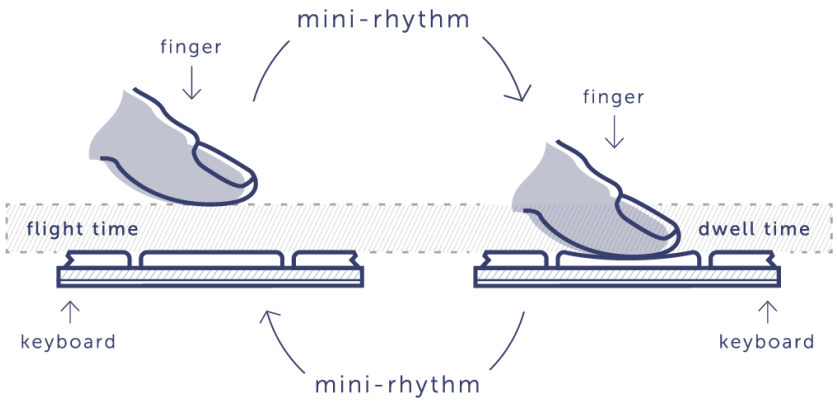
6

-So imagine this. Each user has many devices. Each device has many sensor. Each sensor generates data. And you have many users, thousands, millions...
-So you organize data lake. And of course, it is not bad (assuming that you make it correctly, there was a lot of discussion about it during this conference). But (there is always a but word) what next?



-We have moved from one AI keyword – Data lake into another AI keyword – Feature engineering.
- By the way, it could be quite effective recipe for a good AI presentation – move from one AI keyword to another – it can quarantee that each auditor will find something for himself. :D.
- Ok, to the facts. Raw data from devices is really far from data which could be used for modeling. Timestamps of key pressing times, timestamp of mouse location, number of keystrokes,… we have to organize them.

keyboard feature.



8

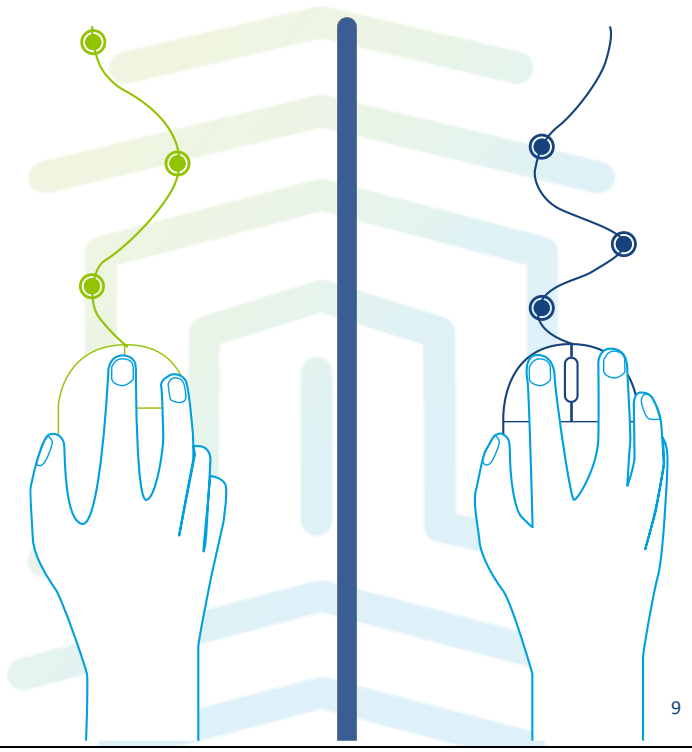
- Keyboard feature example. It can be based on key down and key up events rhythm. Let's take a look at pianists. We can choose several of them and one song. Each of them will play it differently, in his own style.
-The same thing is with the rhythm of typing. It can be uniquely characterize by some metrics like flight time (time between realising the previous key and pressing the next one) and dwell time (time between pressing and realising the same key).
-These metrics can be extended to group of keys, making verification more precise.

mouse feature 1.

What you can obtain from mouse position sensor?

- capture point position { x, y, t }
- silence info - no user action
- mouse-down (button press), mouse-up (button release), wheel scrolling, movement (action finished with click)
- curvature, velocity (horizontal, vertical, tangential, angular), acceleration (tangential, angular)
- click duration, pause before click, pause time
- drag&drop, mouse movement, point click

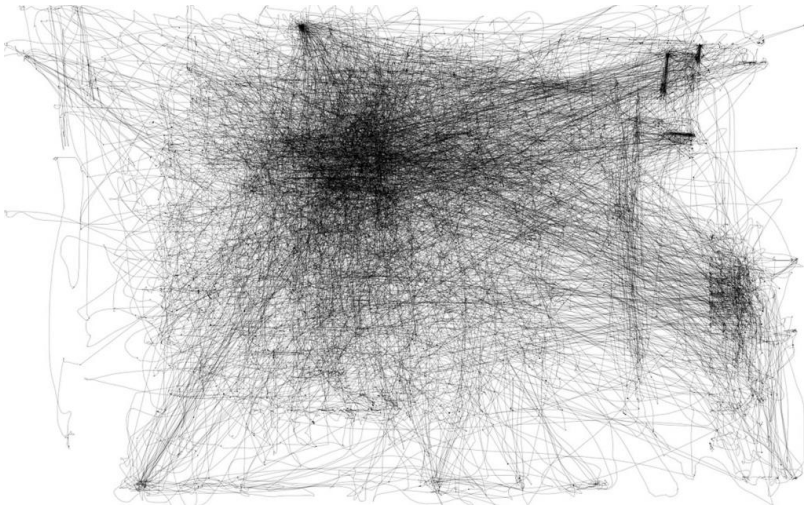
and more



9

- Mouse device gives much more possibilites to characterize user. The simplest info is point position (x and y coordinates with time). Going further, you can define no-action periods, buttons pushing, wheel scrolling, movements (of different types) and different types of velocities and accelarations.
- There is also another level of problem composition – the sampling freuquency – how often you should dump mouse position?

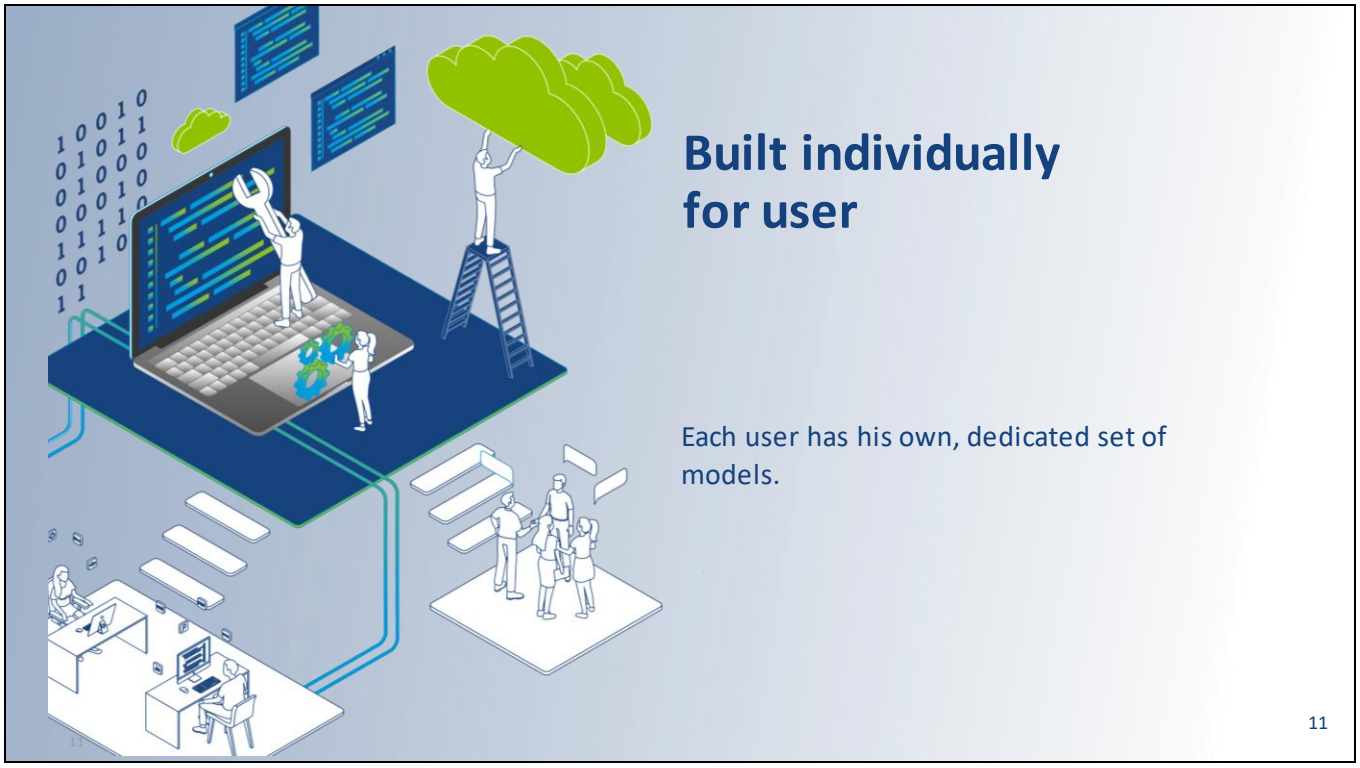
mouse feature 2 - dense areas analysis.



source: Anatoliy Zenkov
https://www.flickr.com/photos/anatoliy_zenkov/4271820522/in/album-72157623274370724/

10

- Another approach to mouse feature building is based on dense areas analysis.
- Instead of bringing down mouse activitites to time series (like acceleration in different timestamp), we can threat them as a heatmap and we can analyze activities areas.
- Recent years brought increasing number of researches about mouse feature. Also with other approach then presented. But they were generally made on quite small sample od 10-20 users. What about big scale learning?



- General approach in behavioural modeling is to build dedicated model for each user. This can give you a good prediction quality.
- But, moreover, each feature (keyboard, mouse) requires own model. So we have to train more than one model for each user. And build another model to link them
- (It could be solved with single ensemble or stacking model, but I do not treat them as single model – you have to train many separates model to get ensemble or stacking model)
- It leads us straighforwardly to training time (which should be low) and further to costs of traning (which also should be low). How to combine it with model quality, especially when you have thousands of users?

Be fast,
be light
and still be precise.

- **Fast** – in case of big scale models training, each second on each single model training is valuable
- **Light** – you have to keep all these models somewhere,
  bigger model -> higher cost
- **Precise** – the essence of ML, no one wants bad efficiency models

DIGITAL FINGERPRINTS

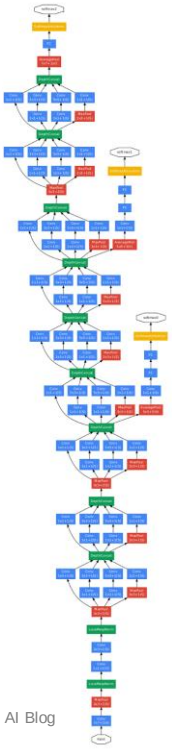
12

- As Digital Fingerprints, we protects more than 1 million users. To train models for this number of users, we have to look for models which are simple enough to be trained fast.
- Because of that, we are focusing more on algorithms based on decision trees, like XGBoost, LightGBM or those based on hyperplane distances (Support Vector Machine, K Nearest Neighborhood) or their ensembles.
- Generally they are faster to train then Neural Networks

Wait, what about Neural Networks?

Every Data Scientist wants to work with Neural Networks. This is, trendy, cutting-edge technology.

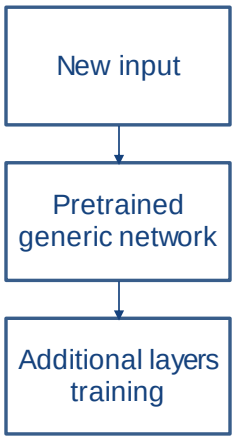
But in the same time it is data and time-consuming, which is a problem in our case.

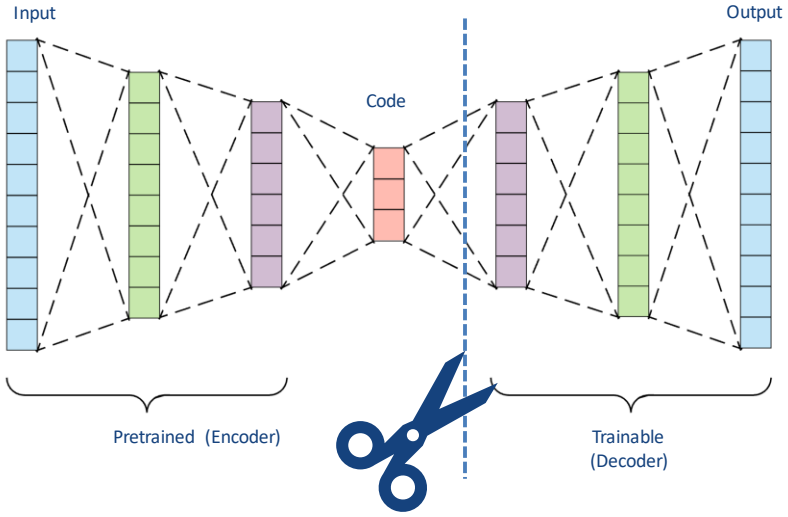


GoogleNet architecture, from Google AI Blog

- Let's spend a moment on Neural Networks. It is top trendy, cutting-edge technology, about which everybody talk (even outside Data Science/ Machine learning circle of people)
- But in the same time it is data-consuming and time-consuming, which is a real problem in Big Scale Model Training
- Let's look at the example of GoogleNet architecture. This example has few years, but gives a nice insight about how complicated NN could be.
- Design of this network required many years of careful experimentation and refinement from initial versions of convolutional architectures.

Transfer learning.





14

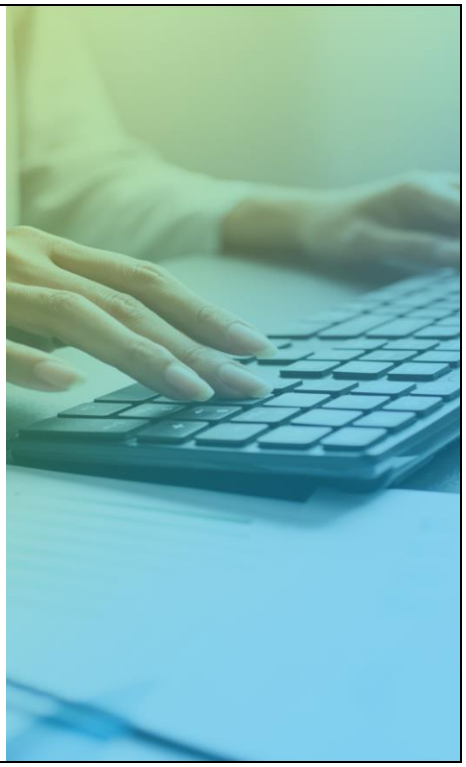
- Transfer learning could help. This machine learning method can significantly reduce model training time for average user
- We can train one, universal Neural Network on big sample of different users. Then, for each user separatelty, we can train just few last layers of bigger network (or even just some chosen neurons from these layers)
- This can reduce training time of single model for each user, which makes this approach applicable to large scale model trainings
- Transfer learning, is the most common in Convolutional Neural Network, but it can be applicable to different Neural Networks too

Be more precise - Hyperparameters optimisation

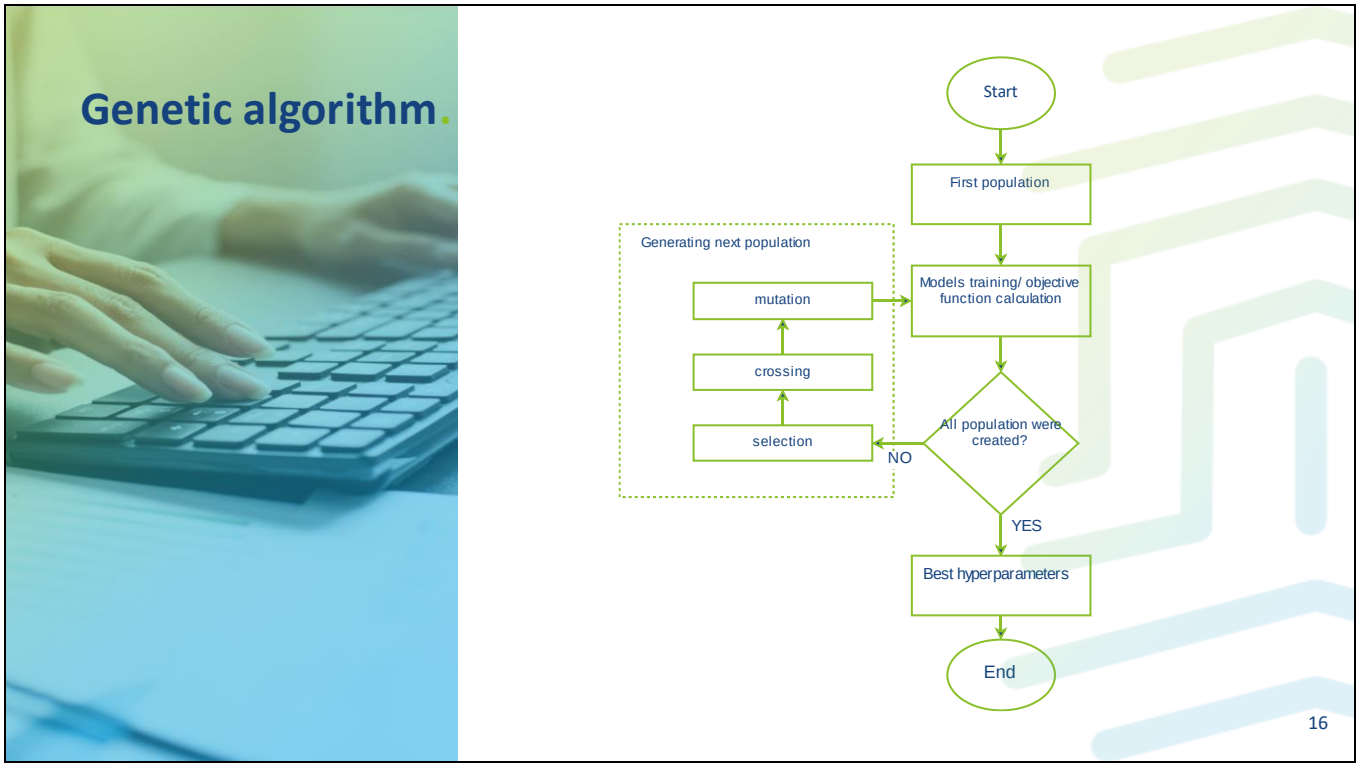
RANDOM SEARCH

GRID SEARCH

GENETIC ALGORITHM



- We can trying to reduce training time (and will be appreciated by your CFO *Chief Financial Officer*) by finally machine learning project should gives good predictions.
- That's the place for hyperparameters optimization, and its methods (from simplest to implement and understand: random search and grid search, to more complicated, like genetic algorithm or bayesian search



- Genetic algorithm - source of inspiration – Darwin theory of natural evolution (here we have another example of biology inspiration)
- General idea is quite simple we are creating first population of models and the we are calculating objective function for each element (model)
- Best elements are transferred to next population, sometimes mutated (by change in one or more hyperparameter), sometimes crossed with other elements
- This way, generation by generation, best elements are passing their gens, which leads us to best model (best set of hyperparameters)

Hyperparameters optimisation and big scale modeling

vs

17

- In big scale model trainings it would be hard to use hyperparameter optimisation in typical way. By which I mean that we can run hyperparameter optimisation for each model.
- So we are looking for alternative solution – some ways to collect users into groups and running hyperparameter optimisation for whole group
- One more time – the amount of models to train doesn't allow to use machine learning techniques in normal form. It forces us to look for an alternative solution (which is really pleasant job :D)

User segmentation.

FEATURE:

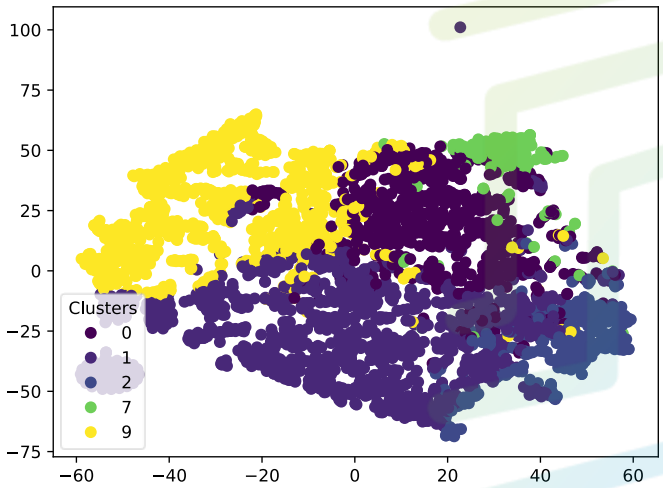
keyboard statistical data

DIMENSION REDUCTION:

t-Distributed Stochastic Neighbor Embedding (t-SNE)

SEGMENTATION:

Gaussian mixture model (GMM)



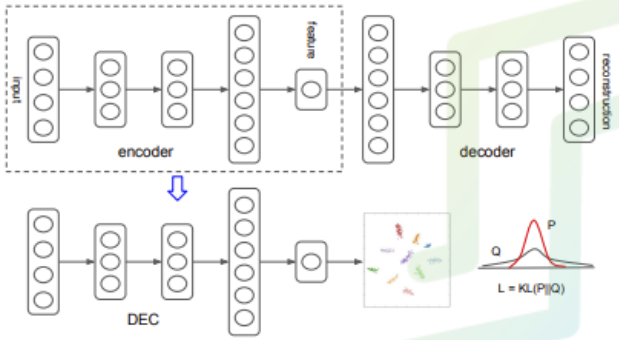
DIGITAL FINGERPRINTS

18

- Here is an example of segmentation made with Gaussian mixture model on one of keyboard feauture developing in our company. We've called it keyboard statistical data
- Dimensionality of this feature was reduced using t-SNE method
- What we finally get is quite distinctive segmention of users, which could be later used to to run hyperparameter optimisation on bunch of users

User segmentation methods.

- DEC (DEEP CLUSTERING EMBEDDING – HYBRID OF AUTOENCODER AND KMEANS)
- DBSCAN
- GAUSSIAN MIXTURES MODELS



source: https://medium.com/@xiaosean5408/dec%E7%B0%A1%E4%BB%8B-unsupervised-deep-embedding-for-clustering-analysis-2b4e7e65e4bf

DIGITAL FINGERPRINTS

19

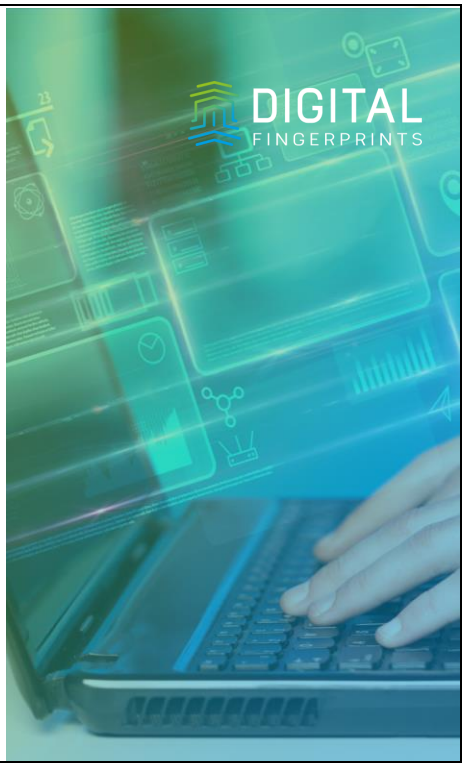
- The list of effective segmentation method in behavioural biometry is longer than Gaussian Mixture Model.
- From the experience of my team I can say that Deep Clustering Embeding based on Autoencoder and Kmeans or DBScan are also approaches worth to try
- An

Thanks for your attention :)

Tomasz Bąk

Head of Machine Learning team in
Digital Fingerprint

tomaszbak@fingerprint.digital



DIGITAL FINGERPRINTS

Mobile – on the way.

We are working on a development base (SDK) for Android that will enable us to collect data.

In coming weeks, we are starting the feature extraction process for mobile data.

21