A REPORT

ON

**FACIAL DEMOGRAPHICS**

BY

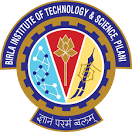
|  |  |
| --- | --- |
| Jash Shah | 2018A8PS0507P |
| Ashutosh Sharma | 2018B3A70928P |
| Vikas Sheoran | 2018B3A70847H |
| Shreyansh Joshi | 2018A7PS0097G |

AT



**Silver Touch Technologies, Ahmedabad**

A Practice School-I station of



**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE,**

**PILANI (Rajasthan)**

**June 2020**

A REPORT

ON

**FACIAL DEMOGRAPHICS**

BY

|  |  |  |
| --- | --- | --- |
| Shreyansh Joshi | 2018A7PS0097G | Computer Science |
| Ashutosh Sharma | 2018B3A70928P | Computer Science &  Economics |
| Vikas Sheoran | 2018B3A70847H | Computer Science &  Economics |
| Jash Shah | 2018A8PS0507P | Electronics & Instrumentation |

Prepared in partial fulfillment of the

Practice School-1 Course No.

BITS F221

AT

**Silver Touch Technologies, Ahmedabad**

A Practice School-I station of

**BIRLA INSTITUTE OF TECHNOLOGY & SCIENCE, PILANI**

**June 2020**

# **Acknowledgements**

The PS-I project has been a great opportunity to work on real world problems and research for optimal and professional solutions.

Firstly, I express my deepest gratitude to Dr.S.Gurunarayanan Sir, PSD Head, for managing this wonderful initiative of Practice School and for providing us with this great opportunity to learn how the corporate world works.

I would like to thank Mr. Adarsh Parikh, Associate Vice President and Training and Development Head at Silvertouch Technologies, for providing us with an excellent opportunity to intern with them and gain exposure to work and Professional ethics.

It is my radiant sentiment to place on record my best regards and deepest sense of gratitude to Mrs/Ms. Tanisha Bhayani, AI/ML engineer at Silvertouch and our Project Manager and Mrs./Ms. Akanksha Goplani, HR at Silvertouch and our project coordinator, for their mindful guidance which was extremely valuable for the project, both theoretically and practically.

I would also use this opportunity to extend my deepest gratitude and special thanks to our PS faculty, Dr. Aruna Malapati , Associate Professor (CSIS), BITS-Pilani, Hyderabad Campus, who, despite being extraordinarily busy with her/his duties, took time to hear each issue, guide and keep me on the correct path throughout this time.

Lastly, I would like to extend my thanks to my project team members who have been highly supportive. The project is a collective effort of all the members, and it would not have been possible without them.

**BIRLA INSTITUTE OF TECHNOLOGY AND SCIENCE, PILANI**

**(RAJASTHAN)**

**Practice School Division**

**Station:** Silver Touch Technologies Ltd **Centre :** Ahmedabad

**Duration:** 6 Weeks **Date of Start:** 18th May

**Date of Submission :** 27th June

**Title of the Project:** Facial Demographics

**ID No./Name(s)/ Discipline(s)/of the student(s):**

|  |  |  |
| --- | --- | --- |
| Shreyansh Joshi | 2018A7PS0097G | Computer Science |
| Ashutosh Sharma | 2018B3A70928P | Computer Science & Economics |
| Vikas Sheoran | 2018B3A70847H | Computer Science & Economics |
| Jash Shah | 2018A8PS0507P | Electronics & Instrumentation |

**Name(s) and designation(s) of the expert(s):**

Tanisha Bhayani, AI - ML Engineer, SilverTouch Technologies Ltd.

Akanksha Goplani , HR at SilverTouch Technologies Ltd.

**Name(s) of the PS Faculty:**

Dr Aruna Malapati, Associate Professor (CSIS), BITS-Pilani, Hyderabad Campus

**Key Words :** Facial demographics, SVMs, CNNs, feature extraction

**Project Areas :** Machine Learning, Deep Learning and their applications in daily life

**Abstract:** The project Facial Demographics aims to take Computer Vision to a whole new level by creating intelligent machines that are able to predict the gender and age of a person, just by looking at a person’s face image. This report throws light on the progress so far, showcasing various techniques used to complete the project.

# **Table of Contents**

**Acknowledgements..............................................................................................................i**

**Abstract................................................................................................................................ii**

**List of Illustrations..............................................................................................................iv**

1. Introduction 1

2. Problem Definition 2

3. Literature Review 4

4. Methodologies

[4.1 Exploratory Data Analysis (EDA)](#_i22wibqizlxh) 6

[4.2 Convolutional Neural Networks (CNNs)](#_xzn0i4skz2j) 8

[4.3 Transfer Learning](#_y3qek4heoyiu) 10

[4.4 Classical ML Techniques](#_p0lf9kliokca) 13

5. Future Works 15

6. Conclusions 15

**Appendices 16**

**References 18**

# 

# **List of Illustrations**

1. **Table I:** List of Research Papers reviewed………………………………………....4
2. **Figure 1:** Histogram of the dataset ( Frequency vs Age) ……………………….6
3. **Figure 2:** Descriptive Statistics of the dataset for the male and female ……………7
4. **Figure 3:** Finding unlabeled, corrupt images and removing them ………………....7
5. **Figure 4:** CNN Architecture ……………………………………………………….9
6. **Figure 5:** Feature Map in Convolutional Neural Networks (CNN) …………….....11
7. **Figure 6:** Transfer Learning Architecture ………………………………………....12
8. **Figure 7:** Use of variations of Gabor filter on different images…………………...14
9. **Table II:** List of abbreviations used…………………………………………..…...16
10. **Figure 8:** Performance vs Amt. of data……………………………………….…...17

# 

# **1. Introduction**

Silver Touch Technologies is a leading and globally accepted IT solution provider and currently at the forefront of digital transformation & emerging technologies to serve the customers across the world.

Silver Touch Technologies Limited was founded in February 1995 with a view to deliver clients with the Information Technology Solutions in the fields of system integration and software services. Now it is into multiple areas like system integration and software development. The company is providing all end-to-end Information and Communication Technology (ICT) solutions to its clients under a single umbrella. It has carried out several national and international joint-ventures which have enabled the company to expand its geographical reach and create a diverse portfolio of clients. The company has executed various projects in India and other countries including USA, UK, France, Australia, Middle East and Germany.

As a part of the IT services provided by Silver Touch, we have been given an AI -ML project. The project is titled ‘Facial Demographics’ . We have to design a model for facial analysis which can estimate age and gender of a given image.

The objectives of the project include:

* To create a model to predict gender from a face image.
* To create a model to classify the face image in one of the classes of the age.
* To create a model to estimate exact integer age from the face image.
* Create a script (and a service) so that the model could be deployed in real life.

The dataset we are using for training our model is the **UTKFace** dataset (aligned and cropped). It is a large-scale face dataset consisting of 20,708 face images. All images are 200 x 200 with annotations of age (0-116), gender (male, female), and ethnicity (White, Black, Asian, Indian, and Others). The images cover large variation in pose, facial expression, illumination, occlusion, resolution, etc.

The report is divided into five sections. First section defines the problem statement in detail and second section gives a gist of literature reviewed for the project. The third section explains in detail the methodology used and is further divided into four sub sections namely EDA, CNN, TL and ML. Fourth section talks about the future work and Fifth section concludes the report.

# **2. Problem Definition**

One of the hottest fields today in Computer Science, Human-Computer Interaction and HMI fields, has the problem of designing an intelligent machine that can estimate the age and gender of a person on its own, using face images, using well known concepts in artificial intelligence. Many big companies around the globe, such as IBM, Google today, are investing in this field. The primary reasons for the same are -

* The boom in image data in recent times
* The boom in computational power in recent times, allowing for better training of deep learning models
* The fact that we can make a machine decipher certain characteristics of humans just by looking from its face, is in itself very intriguing .

The ability of a machine to figure out a person’s gender and age (or even race for that matter), could prove to be a very useful aspect in real life. Since age and gender are 2 crucial factors in identifying a person by face, the model could be deployed in real life to find lost persons. So, the model could take as input race, age and gender of the lost person (given by someone who wants to find him/her) and the machine would then scan whichever face it gets, will check if the predicted age, gender and race match with the given information. That way AI powered devices could help find a missing person in real life.

This technology could also be applied to improve the vision of the humanoid robots and automated systems

This is precisely the problem the project “Facial Demographics” deals with. It aims to analyze images, their distribution and return information on age, gender for each input face image, based on its facial characteristics. The created model, first extracts features (based on facial characteristics) and then uses those extracted high-dimensional features, to make predictions. Some of these features could be wrinkles, hair color, presence of facial hair, etc.

This report contains 3 types of models, that were created for the same purpose, namely -

* CNN
* Transfer Learning
* Classical Machine Learning model

Each model takes as input an image of fixed size (varying from model to model) and outputs the required class/value.

# **3. Literature Review**

We began our work by reviewing literature available on the problem. The research papers provided us valuable insight in the domain of ‘Facial Demographics’. The research papers that we went through:

*Table I. List of Research Papers reviewed*

|  |  |
| --- | --- |
| Research Paper | Author(s) |
| Gender Classification Techniques: A Review | Preeti Rai and Pritee Khanna |
| Face Recognition Performance: Role of Demographic Information | Brendan F. Klare, Mark J. Burge, Joshua C. Klontz, Richard W. Vorder Bruegge, Anik K. Jain |
| Face Recognition and Age Estimation implications of Changes in Facial Features: A Critical Review Study | Rasha R. Atallah, Amirrudin Kamsin, Maizatul A. Ismail, Sherin A. Abdelrahman, Saber Zerdoumi |
| Age estimation via face images: A Survey | Raphael Angulu |
| Convolutional Neural Networks for Age and Gender Classification | Ari Ekmekji |
| Efficient facial representations for age, gender and identity recognition in organizing photo albums using multi-output ConvNet | Andrey V. Savchenko |

These papers talked about various strategies such as different ways of feature extraction such as Gabor filters, LBP, etc. that can be used with classical ML models. The papers also showcased the various CNN architectures used to solve the problem. They compared trainable and non-trainable algorithms’ performance with respect to the demographics of individuals and how composition of the demographic information in the dataset influences the performance.

Papers also pointed out the importance of facial demographics for many applications such as facial recognition, age estimation etc. in numerous areas such as security, law enforcement, biometrics, forensics etc. and the challenges involved such as pose, illumination, expression, aging.

# **4. Methodologies**

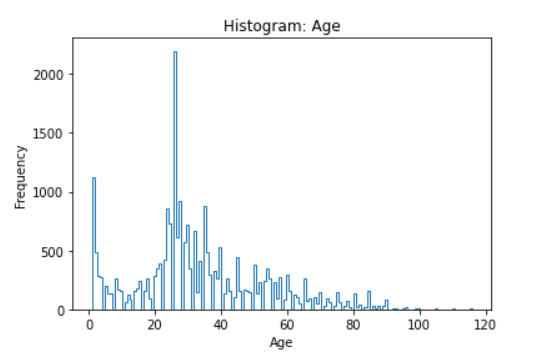
This section of the report, describes in detail various techniques and models that have been created in the project. That is followed by the results obtained by different models.

**Note:** All the code was written in Python and Keras Framework was used as an API which is in turn built on top of Tensorflow.

## **4.1 Exploratory Data Analysis (EDA)**

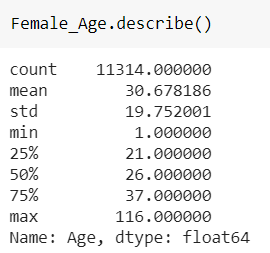
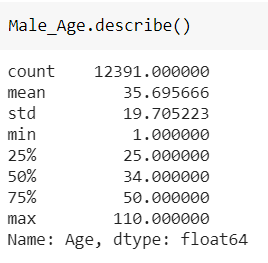
EDA refers to the critical process of analyzing data sets so as to discover patterns, spot anomalies, understand the data distribution , summarize their main characteristics with the help of summary statistics and graphical/visual methods.

It was done to understand the data first and try to gather as many insights from it. EDA helped us in making sense of data in hand, before we started with our models. Some of the EDA techniques we used, and the interpretations are as follows:



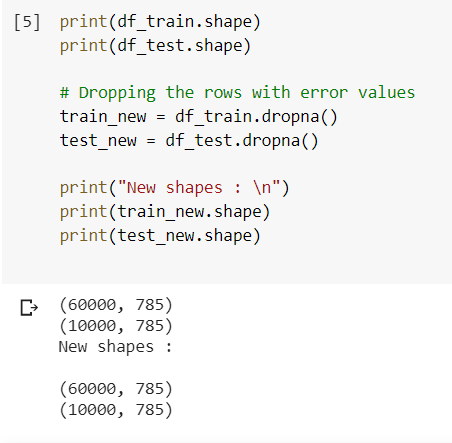
*Fig 1 : Histogram of the dataset ( Frequency vs Age)*

Fig 1. shows that the dataset contains relatively more images with ages between 25-30 and age 1 whereas there are too few images with ages > 80 and ages between 5-15. This to say, tells us that the dataset unbalanced. Now, if we use all images as it is for training, there are much more chances that the ages 25-30 may overfit. So, this tells us that we need to do some preprocessing before using the data to train. We have two options, either we remove the images from ages which have more or we increase the images for ages with lesser (by data augmenting).



*Fig 2. Descriptive Statistics of the dataset for the male and female*

Fig 2. shows the stats of the dataset, just at a glance we can know the number of images in the dataset gender wise, the mean, variance and the quartiles of the ages gender wise. This shows us the distribution of our data and this information comes handy during preprocessing of images.



*Fig 3. Finding unlabeled, corrupt images and removing them*

Having unlabeled images, corrupt images can cause problems in training the model, so we have to get rid of such images. Fig 3. shows this method of cleaning the data. We have made sure that there are no unlabeled images (NaN values essentially).

Some of the other techniques we used were cross-labels mean, variance, histograms, PMFs, IQR, Pie charts, Scatter plots, Box plots, Bee swarm plots, Violin plots, Count Plots, etc.

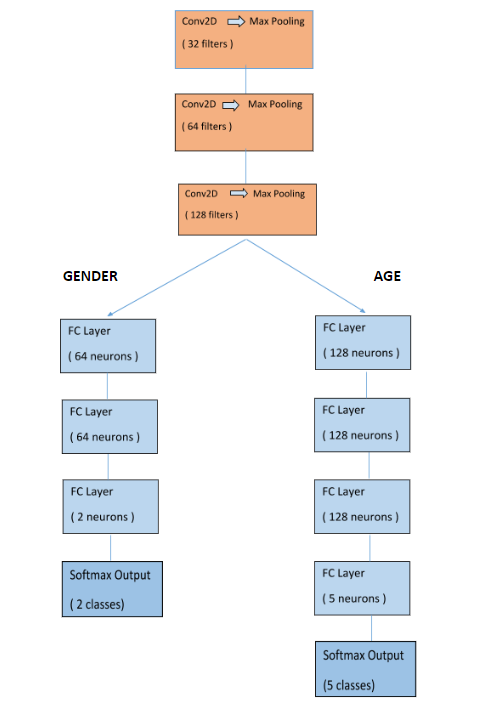
## **4.2 Convolutional Neural Networks (CNNs)**

Images are everywhere around us these days. In the field of DL, when it comes to dealing with images, the most easy to use, comprehensive and ubiquitous way is using CNNs.

Typically, CNNs have 2 main types of layers - **Conv layers**, that are the real feature extractors and **Max Pool** layers, that help in reducing dimensionality, i.e down sampling of features that have been extracted by Conv layers before.

The architecture that was used for gender/age classification tasks is described below -

* Basic idea was to have a block of convolutional and pooling layers, that help the model learn features and representations, followed by a set of FC layers, which are used to classify.
* My model took an input image of size (198,198,3). It was a RGB image.
* To predict both age and gender at once, we used a multi-output classification technique. This involved having a common part of feature extraction for both age and gender, but separate FC layers. The branching point is just after features have been extracted for both. The feature extraction part was kept common for both age and gender classification, because as per researchers, features are more or less along the same lines for predicting age and gender.

****

*Fig 4. CNN Architecture*

* In both these architectures, L2 regularization was used in every FC layer and Dropout (0.3) after every 2 FC layers.
* ‘Learning Rate Scheduler’ of Keras was used to periodically reduce the LR. Initial LR= 0.004 and it halved every 25 epochs.
* Loss functions and metric for both age and gender classification were “categorical cross-entropy” and “accuracy” , respectively.
* BatchNormalization was used after all layers (Conv & FC) in both age and gender classification to normalize the output coming from them, for faster training. It also reduces what is called “covariate shift”.

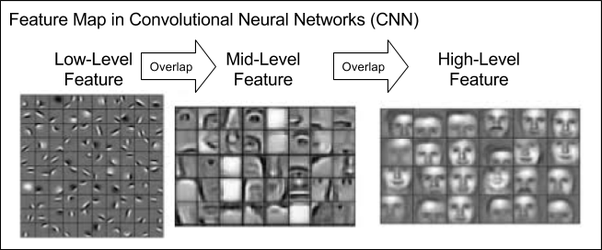
### **Results Obtained:**

Our CNN model, made from scratch was able to achieve a classification accuracy of 85 % on the test set for Age and 94-95 % on the test set for Gender.

For training sets, accuracy was about 96-97 % for both age and gender.

## **4.3 Transfer Learning**

One of the challenges we face in computer vision is the dimensions of feature input. A low resolution picture say, a 200 x 200 x 3 in our case will have 1,20,000 features for each data point and 1,20,001 parameters to train even for a Logistic Classifier which will give us limited performance. If we try a deeper fully connected system we could even have billions of parameters to train! Which will demand a huge amount of data and high computational power to provide decent performance. So we look for another solution and Convolutional Neural Networks(CNN) comes into the picture which attempts to look up for patterns in the image. In the first few layers of CNNs the network using convolutions can identify basic features like the edges and corners, but we can then pass these patterns down through our network and start recognizing more complex features as we go deeper. This property makes CNNs really good at identifying objects in images.

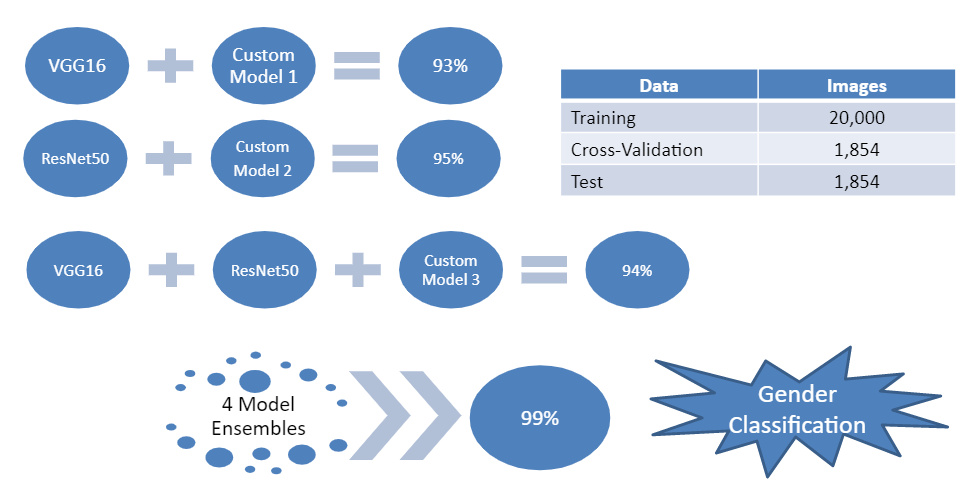


*Fig 5 : Feature Map in Convolutional Neural Networks (CNN)*

We leverage this property in Transfer Learning. We can utilize earlier layers of pre-trained CNN models to detect low-level features or even some mid-level features and save millions of parameters and boost performance by training a model on top of those matured features on limited data.

We utilized the **feature extraction** layers of Oxford VGGFace which is based on **VGG16** architecture by freezing the weights of the ConvNet part. We did a forward pass of the complete dataset to extract features and trained a custom fully connected model on top of it.

We repeated the same for **ResNet50** architecture and trained a different custom model on top of it. Then, we extracted features both from VGG16 and ResNet50 and did **feature fusion** and trained a model on top of that. The accuracy and train, test and cross validation split size for hold-out cross validation are as follow:



*Fig 6: Transfer Learning Architecture*

We also trained a transfer learning model for Age prediction as a real number (regression). The pre-trained model used for the same is VGG16 with VGGFace weights. The model we trained achieved the un-biased mean absolute error of 4.8 as of now against 4.1 as the current benchmark.

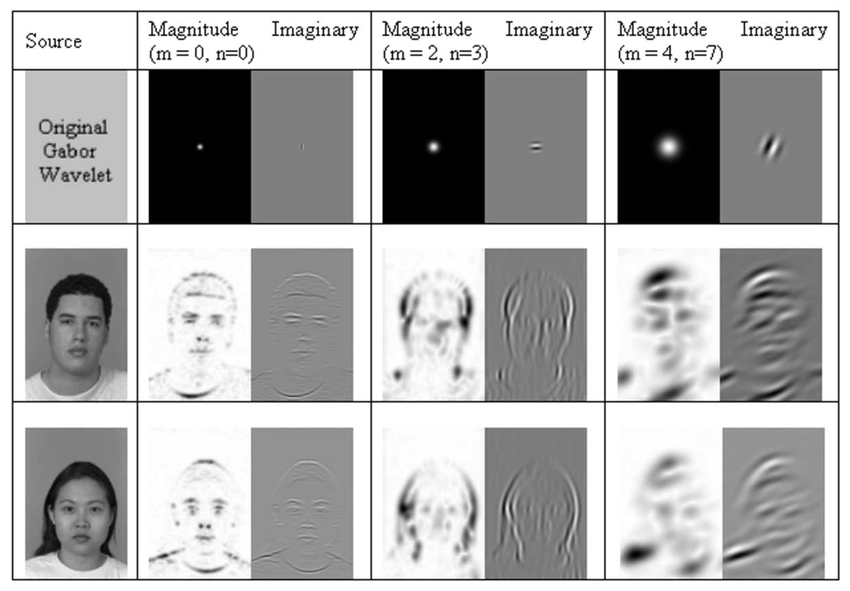
A working prototype has also been made where one can upload a custom photo and make predictions. We believe this could be credited to various Batch Normalization and Dropout layers we implemented.

## **4.4 Classical ML Techniques**

Classical ML models are primitive ways of training a model and involve an extra step of manual feature extraction. This, some time, makes it a better and reliable option while working with small data. Also, at times, a better feature engineering algorithm can increase the overall accuracy of the model. Thus, Classical ML models were studied and SVM was chosen due to its relative high stability in bias-variance trade-off.

Support vector machines (SVMs) are formulated to solve a classical two class pattern recognition problem. We adapt SVM to face recognition by modifying the interpretation of the output of an SVM classifier and devising a representation of facial images that is concordant with a two - class problem.

In the model, the initial feature extraction model was divided into three phases : Image Pre-processing, model free and model-based tracking algorithms. The first stage included image cropping and augmentation stages. Face detection was covered in the second block wherein the color information was exploited to limit the search region. Canny Edge Detection and Histogram Gradient Equalization were used based on the OpenCV framework. This model – free methods are general methods and don’t need any kind of training. The last step for the proposed project involved Gaussian, Gabor and Sobel filters to extract features like Wrinkles, texture, shape and eyes.



*Fig 7 : Use of variations of Gabor filter on different images*

.

### These features were then operated upon the SVM using rbf and polynomial kernels separately and ensembled together. Also, another model was created involving LBP and regularization using LDA and PCA, and these two models were also ensembled at a later stage. A 5-fold Cross-Validation aided the SVC’s accuracy by accounting for the outliers.

### **Results Obtained :**

The current model provided an accuracy of about 94% on the Gender model. In case of age prediction, the model gave a range ( of 5 years ) of age with about 96.7% accuracy.

# **5. Future Works**

For the coming weeks we plan to work on improving the model and extending its capabilities :

1. Hyper parameter tuning and improving accuracy.
2. Creating an age estimation (regression) model.

3. Creating a pipeline for the project to be deployed in real life.

# 

# **6. Conclusion**

The project involves extensive analysis and classification of features. The initial Literature review helped us a lot in understanding the methodology and concepts involved in the project. Again, it is important to reiterate that the quality of the model is highly dependent on the quality of the data . Hence, we performed the EDA process in which we not only analyzed local and global features, but also learnt about the data set and its distribution . This helped us in determining the class weights.

Later on different models were created and benchmarked for their accuracy. We even created predictive models where the user can input an image of their choice. Interestingly, we observed the models are robust to covariate-shift and make astonishing predictions on human face artwork-portraits!

# **Appendices**

## 

## **Appendix A - List of short forms used**

*Table II : List of abbreviations used.*

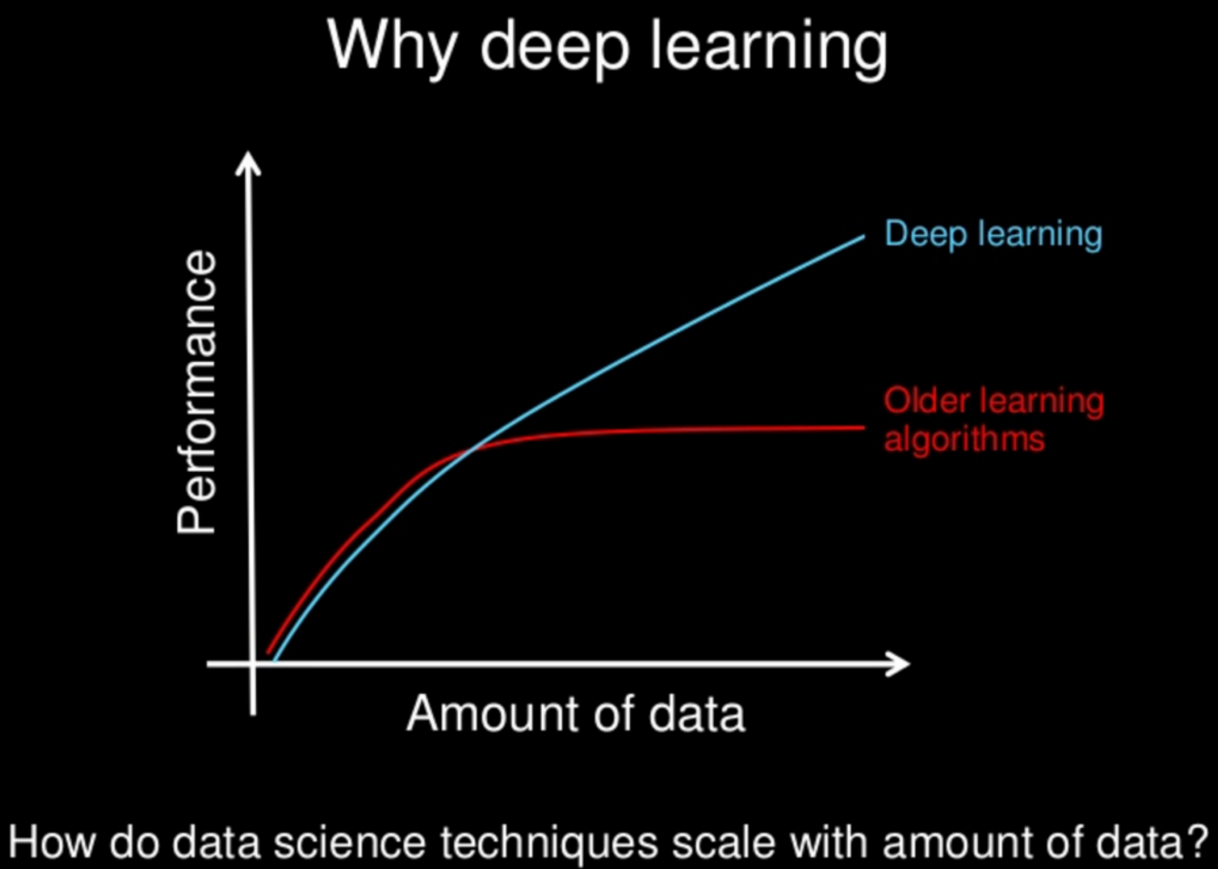
|  |  |
| --- | --- |
| **Short Form (in text)** | **Full Form** |
| EDA | Exploratory Data Analysis |
| CNN | Convolutional Neural Networks |
| ML | Machine Learning |
| DL | Deep Learning |
| SVM (C) | Support Vector Machines (Classifiers) |
| HMI | Human Machine Interaction |
| LDA | Linear Discriminant Analysis |
| LBP | Local Binary Patterns |
| PCA | Principal Component Analysis |

## 

## 

## **Appendix B - Classical ML vs DL models**

* Pros of DL
  + Best in class performance.
  + Scales effectively with data
  + No need for feature engineering
  + Transferable , adaptable and robust to outliers to some extent
* Pros of ML
  + Work better and faster on small dataset
  + Financially cheap
  + Computationally less expensive
  + Easier to interpret and debug, since all features are created by the programmer and then fed into the model



*Fig 8 : Performance vs Amt. of data*

# **References**

1. About Our Company - Silver Touch Technologies. (n.d.). Retrieved from <https://www.silvertouch.com/about-us/>
2. Alexander, Rosebrock, A., Peng, Sam, Abraham, N., Wilf, . . . Seed, L. (2020, April 18). Keras: Multiple outputs and multiple losses. Retrieved June 11, 2020, from <https://www.pyimagesearch.com/2018/06/04/keras-multiple-outputs-and-multiple-losses/>
3. Bansari, S. (2019, April 30). Introduction to how CNNs Work. Retrieved June 10, 2020, from <https://medium.com/datadriveninvestor/introduction-to-how-cnns-work-77e0e4cde99b>
4. Chollet, F. (n.d.). The Keras Blog. Retrieved June 11, 2020, from <https://blog.keras.io/building-powerful-image-classification-models-using-very-little-data.html>
5. Rcmalli. (2020, March 19). Rcmalli/keras-vggface. Retrieved June 11, 2020, from <https://github.com/rcmalli/keras-vggface>
6. Seif, G. (2019, May 04). Deep Learning vs Classical Machine Learning. Retrieved June 11, 2020, from <https://towardsdatascience.com/deep-learning-vs-classical-machine-learning-9a42c6d48aa>