2019 IEEE Intl Conf on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking (ISPA/BDCloud/SocialCom/SustainCom)

Facial-Based Personality Prediction Models For Estimating Individuals Private Traits

Raad Bin Tareaf*, Seyed Ali Alhosseini*, Christoph Meinel†
Hasso-Plattner-Institute
University of Potsdam, Germany
* {raad.bintareaf, seyedali.alhosseini}@hpi.uni-potsdam.de
† {christoph.meinel}@hpi.de

***Abstract*—The essence of images users post and share at their social media platform is motivated and elucidated by their individual psychological constructs which are designated as personality traits. In this research, we investigate how social media profile pictures differ based on the personality of the users posting them at their social networking sites. In our experiment, we use profile images from Twitter platform whose personality we predicted based on 1.7 million data points. We conducted our analysis on users faces by extracting unique 50 facial features in order to examine the relationship between personality and profile picture. Our results reveal notable variations in profile picture selection between different personality traits. For example, high to openness males have a high surprise emotion in their profile pictures and a low smile value, while high in openness females have a high happiness emotion in their profile pictures and a high smile value. Finally, various machine learning approaches were investigated to test the effectiveness of these facial features in predicting users' psychological traits. Our results show that training personality models on a granularity based on gender gains higher accuracy. To our knowledge, this work is the first attempt of using ensemble learning methods for personality prediction task from users profile picture.**

Keywords: Big Five model, Profile Picture, personality prediction, Facial Features, social media analysis.

I. Introduction

Social networking sites plays an important role of our everyday life and users are now more free to choose from various social networking platfroms. In 2017, more than half of the globe uses the world wide web network with 2.7 billion live social media accounts worldwidely [1] and every user of that social platforms leaves a mark as digital footprint and tend to present themselves in a type of behavior usually determined by their psychological constructs known as personality traits.

Considering the variation and differences between users personality, it has shown to be a valuable factor to rely personalisation services on [2]. Consequently, once online users personality is revealed, it can be used to automatically recommend and personalise advertsments that fits users in all aspects. Extensive research efforts is now studying the associations among personality and behaviors (e.g.,marketing [3] [4], movies [5], health [6], music [7][8], education [9][10]).

The increased attention in identifying online users personality raised the questions on how to automatically obtain individuals personality from thier online fingerprints. The legacy approach is to complete a self-report judgments: answering a questionnaire that is levereged to estimate users personality. For example, a well-known NEO Personality Inventory which contain 20 to 360 personality related questions [11]. The models are derived from the combination of words and human personality which appoint five global factors abbreviated as OCEAN: *Openness to Experience*, *Conscientiousness*, *Extraversion*, *Agreeableness* and *Neuroticism*. However, nowadays questionnaires considered a time consuming and intrusive task.

To overcome the cumbersome of using questionnairs, several researchers have attempted to predict users' personality and interests across different contexts and environments in rapid and cost-effective way. Indeed, they were able to accuretly capture users' personality from digital footprints over different social media platforms. In some cases, the prediction were more precise than evaluation made by users friends or family members as mentioned in [12]. Many studies have managed to successfully build models to predict a wide range of user private traits, these studies used different types of information, ranging from social network attributes and connection to text from posts/tweets [13], likes history [14][15] or profile picture choice [16].

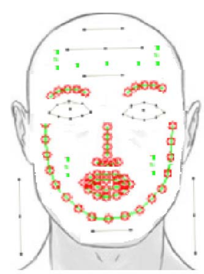

Fig. 1. Visual Representations and Properties for Human Facial Features

With images getting more popularity especially among younger people [19] and recent social networks are focusing on visual content such as Snapchat or Instagram, personality dimentions in this case can be calculated by by running content analysis which is based solely on images as presented in previouse researches [17]. Images, in general, contain multiple variables such as scenes, compositions, colors, emotions, facial presentations, and facial expressions as presented in figure 1 where these properties can be fetched by leveraging

978-1-7281-4328-6/19/$31.00 ©2019 IEEE
DOI 10.1109/ISPA-BDCloud-SustainCom-SocialCom48970.2019.00233

various computer vision algorithms such as [25]. These various descriptions can be leveraged to investigate the difference among users' personality and image uploading behaviour over various types of social networking sites. In this study, we examine all above-mentioned image properties and shows their relationships with personality traits from the Big Five personality model.

This research paper is established as presented in this section. The section II mentions the previous research efforts in the topic of personality prediction from various fingerprints specially from images. Section III discuss the methodology we follows for pre-processing and cleaning our data, we also discuss the process of feature evaluation and selection. Where in section IV, we summarize various examinations of different machine learning algorithms. Lastly, in section V and section VI we conclude latest outcomes with allusion for futuristic progress that surround automating personality detection from profile images.

II. Related Work

The prevalence of social platforms triggered multiple research endeavors in human personality estimation and prediction. In this part, we summarize recent research efforts in predicting individuals personality from appearance. In contrast to traditional ways to calculate users's personality, leveraging social footprints such as images for estimating personality assure simple and fast intuition.

Liu et al. [16] presented a large-scale analysis of profile images and personality at Twitter microblogging platform. They used a wider range of interpretable aesthetic and facial features to capture correlations with the personality that is in line and complement psychological research. They showed that each personality dimension has a different type of profile picture posting behavior. For example, individuals who are high in conscientious or extraversion uses pictures with at least one face and they prefer to present emphatic and positive emotions within their facial expressions. At the evaluation part, they tested the predictive performance of their used features and reported relatively robust precision at the testing samples.

Skowron et al. [20] proposed a unique technique that combines multiple inputs as text, image, meta-features and integrates it out of two different social networking sites which are Instagram and Twitter. They also addressed the issues of dimensionality reduction and noise reduction using subsampling. Furthermore, they used random forest regression for creating a low bias and variance model by averaging regression tree decisions. Their outcomes conclude that the collection of features and the type of the social networking sites, produce variations in final regressor results. The superior outcomes for every personality trait are achieved by blending engineered features derived from jointly social networking sites.

Nie et al. [21] addressed the question of how to estimate social user personality by using the picture choice. They presented personality as a reverberation of individual behaviors on a specific social platform. They divided up the used sample to different sets and then labeled them with various personality dimensions by clustering approach. They introduced low-level features to train and estimate personality scores out of personal photos. The final tests shows the validation of using such a method. They also highlighted the importance of features refinement and design as well for the necessity of enlarging the used samples.

Cristani et al. [22] investigated the rising size of multimedia information users generate and engage online and consider it as a probable contributing factor to our what so-called online appearance. Their work reported tests on the interrelation among users personality and Flickr images. The investigation draws attention to new challenges for in this domain, as detecting visual patterns that gets together with personality traits in an intense direction than regular features. The paper also confers that visual patterns correlates with personality score and can be used to predict personality where also they found that the favourite images users assign in his/her profile can be used eventually to build prediction models to estimate their preserved online personality.

III. Implementation

The following part of the research paper is splitted onto multiple sub-sectionw. The retrievements of the used data for training phase as well as for the testing datasets from both MyPersonality dataset and the crawled Twitter dataset in (III-A). The following sub-section represent the process of feature induction using the *Linguistic Inquiry and Word Count* tool then the usage of *Face++* tool, datasets used to build final models in (III-B) and the process of picking up the most significant features to statistically significant level in (III-C). The process of training various regressions and ensemble algorithms in (III-D), where Section (III-E) conclude the final quality measurements.

A. Datasets

We used two different data sets in our experiments (Facebook and Twitter). For the aim of personality estimation, only a few trusted and ground truth datasets are obtainable. To mention some, MyPersonality's app dataset considered as one of the most advanced and benchmarked dataset for researchers among onlie personality estimation. In our research efforts, we leveraged MyPersonality's dataset as the training data to build the final personality prediction models from textual inputs.

Also, for the task of estimating individuals personality through portrait images, we did not find any trust-able dataset to work with. Therefore, we decided to randomly collect a huge amount of Twitter profiles from Twitter API considering users full tweets history and the url of their profile picture. The collected dataset was not labeled and can not be leveraged to let various algorithms learn to predict personality from images. So we used the Facebook self-reported samples to built personality prediction models to predict personality from texts and then generalise it to predict personality of the second data set (Twitter) from their tweets. Later, we concatenated the predicted Twitter personality from users tweets with their profile picture url. We discussed in more depth and details the validity and the potential of the used learning mechanism and how we propagate the personality scores from text to images samples.

*1) **MyPersonality Facebook Dataset**:* This dataset was a widespreaded Facebook application created by [23] in 2007.

The application let individuals to take a part in diverse psychometric exams, such as the Five-Factor questionnaire which is to some extent similiar to the NEO Personality Inventory from [11]. Nearly 30 percent of the individuals who took a part in the questionnaire agreed to share their public information with the application. MyPersonality application dataset represent almost six million psychometric questionnaire results for four million unique individuals.

In our research, we only leveraged the three main datasets from the application dataset which are:

Demographic Details Table: It represent the demographic information of four million Facebook users with their Facebook ID, gender, age, relationship status,events attended, *Interested In* information, language, number of friends and many other related information.
BIG5 Personality Scores: Holds the whole generated personality scores varying form $[1-5]$ with the information about the survey length each user filled.
Facebook Status Updates: It is the table which contains the whole status updates for the users who was participated in the application. It present a twenty five million status update for individuals.

The studied dataset also have the (Linguistic Inquiry and Word Count) tools labelling for almost 153.617 Facebook individual. These scores were computed by executing the (Linguistic Inquiry and Word Count) application [24] at the users level. The (Linguistic Inquiry and Word Count) analysis tool estimate the varying emotions and thinking styles or social concerns in any given text and save the output in any desired format. Every single annotation which is also refereed as word categories, is demonstrated as a percentage of words for all individual's status posts.

TABLE I. CHARACTERISTICS OF FACEBOOK TRAINING DATASET

Characteristics		
# Samples = # Users		108547
# Male		44844
# Female		63245
# Features		93
# Labels		5
Avg. Age		27
Labels	**Mean**	**Standard Deviation**
Openness	3.8435	0.6759
Conscientiousness	3.4631	0.7358
Extraversion	3.5068	0.8135
Agreeableness	3.5659	0.7070
Neuroticism	2.7334	0.8003

Multiple feature extraction methods (pairwise Pearson product-momentum and gradient boosting trees) has been investigated to capture the best correlated and significant features for each personality trait and then used them to build a sophisticated personalty models that can precisely predicted users personality from their textual inputs. Furthermore, multiple machine learning algorithms as (Support Vector Regressor, XGB Boost and Feed Forward Neural Network) built and used to evaluate the final models as reported in Table II. For more details about all followed methodologies and concepts that generate the final models, please refer to Bin Tareaf et al. [4]

TABLE II. COMPARISON OF (RMSE) OF THE SVR, BOOSTING AND NN MODELS TRAINED ON DIFFERENT FEATURES SETS. BOLD VALUES INDICATE LOWEST ERROR FOR THE ASSOCIATED TRAIT.

Trait	**Baseline**	**SVR**	**Boosting**	**NN**
Neuroticism	0.7963	0.7689	**0.7635**	0.7611
Openness	0.6779	0.6382	**0.6309**	0.6525
Extraversion	0.8079	0.7672	**0.7655**	0.7877
Conscientiousness	0.7381	0.7003	**0.6934**	0.7279
Agreeableness	0.7051	**0.6786**	0.6823	0.7069

*2) **Twitter Random Dataset**:* Using Tweepy python package, we randomly collected 1.7 million data points from Twitter REST API by only selecting public profiles that meet our filter criteria (only one face in the associated picture). The process yielded of more than 600 unique Twitter users (383 males and 227 females). For each user, we collected almost 2800 tweets in average associated with a hyper link for users current profile picture.

TABLE III. CHARACTERISTICS OF TWITTER GENERATED DATASET

Characteristics		
# Samples = # Users		610
# Tweets		1.7 m
# Male		383
# Female		227
# Facial_Features by Face++		50
# Generated Labels		5
Predicted Labels	**Mean**	**Standard Deviation**
Openness	3.6277	0.7553
Conscientiousness	3.3581	0.7819
Extraversion	3.1162	0.7537
Agreeableness	3.3157	0.7932
Neuroticism	2.5691	0.7108

We used the Facebook prediction models that we described before (trained on features extracted by Pearson correlation) to label our current Twitter samples with personality scores. Therefore, we ran our algorithm over all collected twitter samples and eventually the iteration process returned all Big Five personality measurements for all samples as presented in Table III. Furthermore, we used the link of profile picture for each user to feed it directly to a deep learning based method Face++ API [25] to extract all facial features from each user image, resulting in 50 facial features as presented in the appendix in Table VI. Face++ algorithm has approved its superiority in providing an accurate face recognition, demographics and facial presentation. At this point, we created a new data set that contains users Twitter personality scores associated with his/her facial features scores.

B. Samples

The accumulated dataset that we leveraged to train the prediction models has 1.700.000 data points for almost 600 Twitter user. This amount of data allows us to use a decent amount of samples to train the final models. In order to evaluate the final output model, we used a sampling technique which is random split to pull out almost 20 % of the available samples to anticipate them as a final testing samples for our final prediction models.

C. *Feature Selection*

After multiple experiments, we found that if we segregate our data samples based on gender to build personality prediction models it will yield better and accurate results. Therefore, we analysed various males and females facial features associated with their personality scores. For each personality trait, we trained a dedicated model to predict it as described in (see III-D). We investigated various feature sets for each predictor and then decided which facial feature affect the appearance of specific personality traits and how gender as a key factor plays a dominant role when it come to predict personality solely from users profile images.

We used two different approaches to examine and select the appropriate facial features sets for both genders. First method considers the Pearson's Correlation Coefficient and the other method utilizes the Boosted Decision Trees' feature significance. All the above mentioned methodologies in extracting features sets for both genders are discussed in the following two sections. Figure 2 represent the inter-correlation between all personality traits for both females and males Twitter samples, and Figure 3 and Figure 4 represent the inter-correlation between the facial features for both females and males Twitter samples.

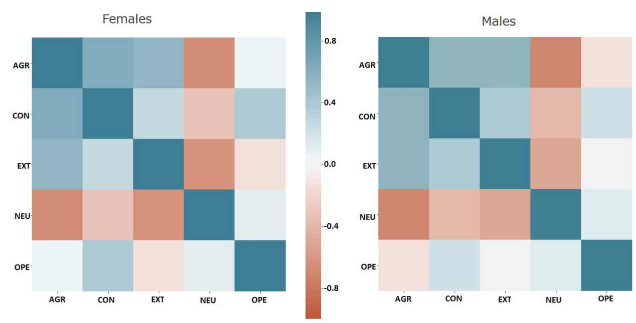

Fig. 2. Pearson Inter-Correlation Coefficient Heatmap between the Big Five Personality Traits Between the Same Set of Users for Both Males & Females Samples. Blue = High Coefficient.

*1) **Pearson Correlation**:* Pearson correlation assumes a linear relationship between two variables. Therefore, we applied the correlation test as pairwise correlation analysis among the 50 facial features and personality records by utilizing the *Pearson product-moment correlation*. Following this methodology, we were able to derive $m = 5$ correlations with the same feature set (because the task is to predict five different personality traits). In order to over come the various comparing issues, we enforced what so called the *Bonferroni correction* to our global significance level of $\alpha = 0.05$ to decide the local significance levels: $\alpha^* = \frac{p}{m} = \frac{0.05}{5} = 0.01$.

In order to comprehend the correlation values and significance levels between the extracted facial features and the five personality traits, we visualise them in a heatmaps as presented in Figure 5 and for females samples in heatmap Figure 6. Typically, psychological variables have a correlational upper bound between 0.3 - 0.4 as presented in [28]. Surely, not all facial features are correlated with the personality scores. To mention some, there are a facial features returned by Face++ algorithm but they have no correlation and significance with the per-

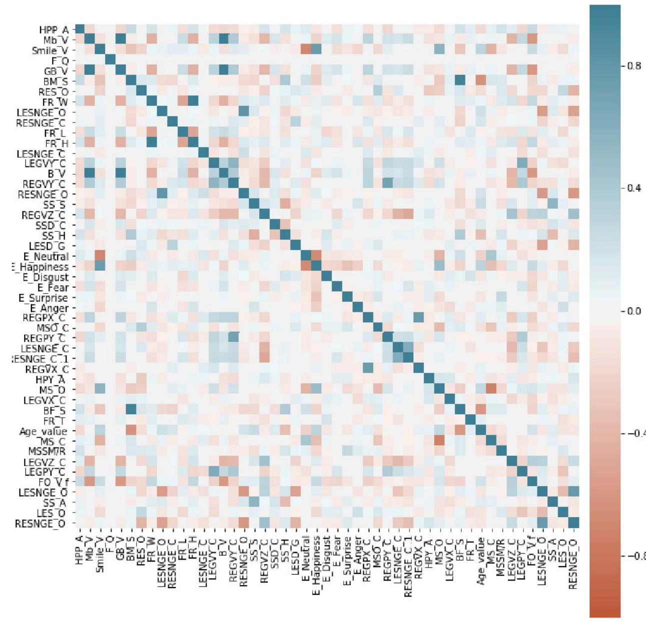

Fig. 3. Pearson Inter-Correlation Coefficient Heatmap Between the Females Facial Features and the Same Set of Users. Features Full Name List is Available in Table VI in the Appendix. Blue = High Coefficient.

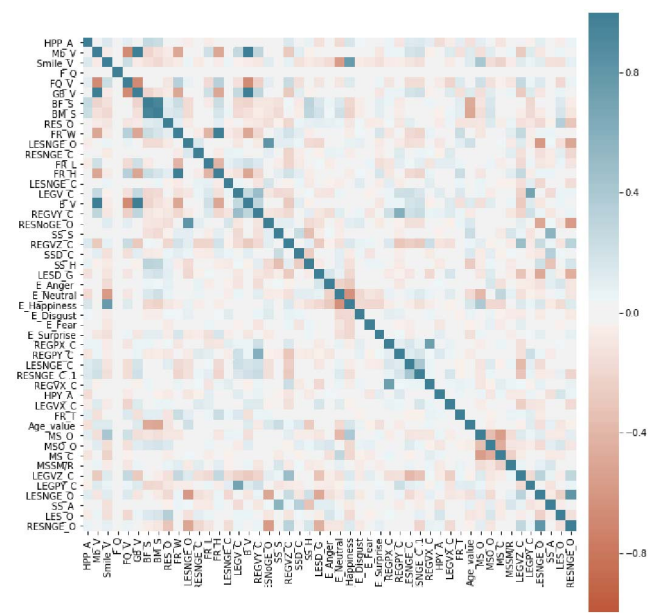

Fig. 4. Pearson Inter-Correlation Coefficient Heatmap Between the Males Facial Features and the Same Set of Users. Features Full Name List is Available in Table VI in the Appendix. Blue = High Coefficient.

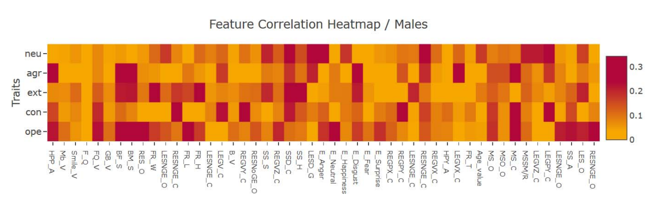

Fig. 5. Pearson Correlation Coefficient Heatmap for Male Samples. Red = high coefficient.

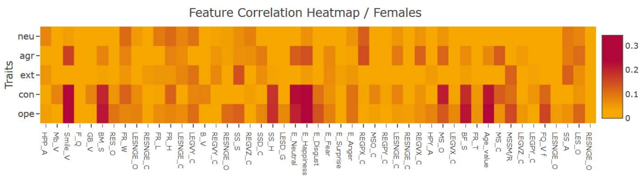

Fig. 6. Pearson Correlation Coefficient Heatmap for Female Samples. Red = high coefficient.

sonality traits for both genders, such as *blurness_threshold*, *gaussianblur_threshold*, *glass_value* or *motionblur_thresold* .

Facial features with overall high relative correlation coefficients for females are e. g. *smile_value*, *skinstatus_health*, *mouthstatus_open* and facial *emotions* in general as happiness, where for males are e. g. *right_eye_status*, *mouthstatus_close* or *skinstatus_darkcircle*. Where correlations for the Neuroticism trait specially among females samples are harder to detect compared for other traits. Interestingly, this conclusion come in line with previous research we conducted which also reported difficulties in predicting Neuroticism trait from normal textual inputs.

To define all influential features for a specific personality trait, we took into consideration only features that shows a significant correlation ($p < \alpha^* = 0.01$) for the studied trait *X*. As a result, we minimized the number of features by choosing features that presented a strong correlations ($|r| > 0.05$). For example, the final feature set for the personality trait *Openness* is selected by choosing the features that significantly correlated with $p < 0.01$.

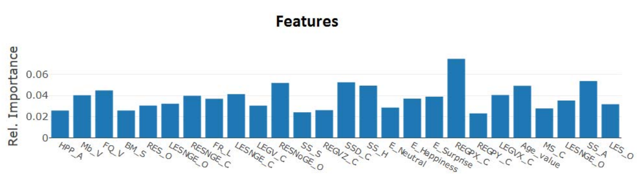

Fig. 7. Significance and relative importances for Male-Agreeableness: The diagram contains all features with a relative importance higher than 0.011 value.

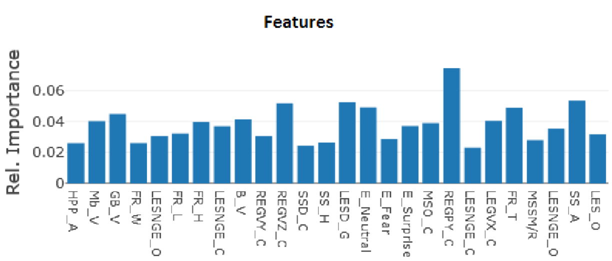

Fig. 8. Significance and relative importances for Female-Agreeableness: The diagram contains all features with a relative importance higher than 0.011 value.

*2) **Gradient Boosting**:* The second methodology we investigated to extract an adequate set of features for each personality dimenstion is to compute the relative importance

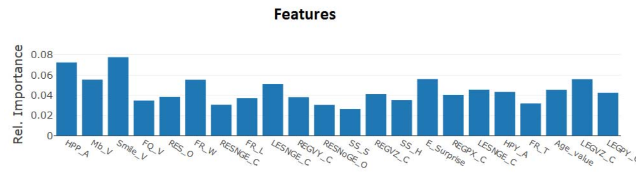

Fig. 9. Significance and relative importances for Male-Conscientiousness: The diagram contains all features with a relative importance higher than 0.011 value.

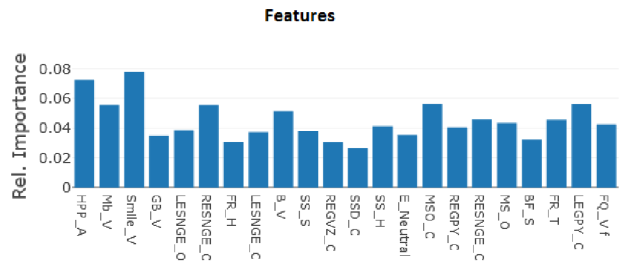

Fig. 10. Significance and relative importances for Female-Conscientiousness: The diagram contains all features with a relative importance higher than 0.011 value.

by using the gradient boosted regression tree [26]. The main concept of gradient boosted regression tree is to learn a prediction model using all available features on the dataset. The final estimator will implicitly have the importance of each feature in forming the final prediction.

As we have limited space in this research paper, we can't present all cases for both males and females with all relative importance features with all personality traits. Therefore, we decided to report only the agreeableness and conscientiousness traits for both males and females samples as presented in Figure 7, 8, 9 and 10 respectively.

The personality trait Extraversion and Neuroticism are most difficult traits to be predicted, specially for females samples, because their feature sets does not contain a decent significant correlated features. Meanwhile, the approach which we consider to further extract features is the gradient boosting method, it managed to define multiple important features for each personality factor. Therefore, from this information we tested the hypothesis of having non-linear relationships among facial features and personality traits that are hard to predict such as Extraversion and Conscientiousness and this what lead us to investigate multiple feature extraction methods with different machine learning algorithms as reported in the next section.

D. Training Machine Learning Algorithms

Considering the fact that the big five personality results are continuous scores between 1 to 5, we investigated the regression algorithms to estimate individuals five personality dimensions. Furthermore, regression algorithms estimate a mapping function from the feature vector to a continuous output variable. Considering our learning samples, we trained three different machine learning algorithms which are support vector regression, Random Forest, Bagging and Adaptive

Boosting. All of these algorithms are characterized in the next part.

The five factor personality model characterize an individual's personality through five personality dimensions, therefore we decided to train five algorithms where we built and train a dedicated model for each personality factor. Every estimator is learned on two trait specific features, defined by two different methodologies for all five personality dimensions (see section III-C).

All four trained estimators need some tuning for the hyperparameters. We used the Mean Squad Error metric for evaluating the final accumulated error. The used metric for evaluation is detailed in section III-E.

*1) **Support Vector Regression**:* Support Verctor Machines represent the input vectors in an infinite dimensional vector space as we introduce in [27]:

$$K_{RBF}(\boldsymbol{x}, \boldsymbol{x'}) = \langle \phi(\boldsymbol{x}), \phi(\boldsymbol{x'}) \rangle \quad (1)$$

$$= exp\left(-\frac{\|\boldsymbol{x} - \boldsymbol{x'}\|^2}{2\sigma^2}\right) \quad (2)$$

SVR calculate the likeness of two feature vectors $\boldsymbol{x}$ and $\boldsymbol{x'}$ in the input space. The Kernel $K_{RBF}(\boldsymbol{x}, \boldsymbol{x'})$ is large if the euclidean distance between the two feature vectors $\|\boldsymbol{x} - \boldsymbol{x'}\|$ is small. The *rbf*-kernel has one free parameter σ. Together with the regularization parameter C of SVR two hyper-parameters can thus be optimised in the learning phase.

*2) **Random Forest**:* Random Forest is an ensemble technique capable of performing both regression and classification tasks based on multiple decision trees. To create a decision tree, the data is mapped into an n-dimensional space where each dimension stands for one feature. The algorithm then tries to define a decision boundary and divides the dataset into two non-overlapping partitions. This is continued until a remaining group can be perfectly separated which could result in very small buckets and overfitting having a very low bias but a high variance. To mitigate overfitting, a stopping criteria could be introduced like a maximum tree size assigning the majority value as a result to a bucket (tree pruning).

*3) **Bagging**:* Bagging also known as Bootstrap Aggregation and it utilized when we intend to minimize the variance of a decision tree. The main concept is to make several subsets of data from the learning sets decided randomly with replacement (where random forest takes the random selection of features instead of utilizing the whole features to enlarge trees). Every group of the data is utilized to learn their own decision trees. Consequently, it generate an ensemble of various models. Therefore, instead of using single decision tree, an average of all the estimators from several trees are utilized to have a better final results.

*4) **AdaBoost**:* Known as Adaptive Boosting and used mainly to to let the model learns from mistakes by increasing the weight of misclassified data points. The learning process starts by initialize the weights of data points, then train a decision tree for each trait. Later, we calculate the weighted error rate of the decision trees known as *e*. We calculate the decision trees weight in the ensemble and then update weights of wrongly classified points. We repeat the mentioned steps until we make a final prediction. Obviously, the tree with higher weight will have more influence at the final decision.

*E. **Quality measures***

Multiple evaluation metrics are available for data scientists to evaluate the correctness of thier final prediction models. As a result, we decided to assess the final prediction models using the mean squad error metric. It calculates the variation between personality predicted scores against the real actual self-reported scores. The metric is represented with the following formula:

$$MSE = \frac{\sum_{t=1}^{n} (y_{t,act} - y_{t,pred})^2}{n} \quad (3)$$

IV. RESULTS

In this section, we visualize the effectiveness of the four trained algorithms' among multiple feature sets. Finally, we evaluate the performance of every algorithm we trained. Both feature extraction approaches we introduced before produce a very similar outputs based on the (MSE), which also perform better comparing it to the baseline for all personality dimensions.

As Table IV points out, the Random Forest algorithm produced the most accurate results when trained on trait specific feature resulting in a lower errors for four personality traits concerning the males samples. SVR performed very well (confirms several previous researches in this area) for the prediction task, interestingly, it were able to outperform Random Forest prediction for the Neuroticism personality trait, we report the behaviour of our testing male data set in Figure 11 (only few traits diagrams are reported for brevity and space sake). On this dataset, Extraversion were the easiest between all personality traits to predict where Openness was the hardest for males samples. Despite the small sample size, Bagging and AdaBoost were the worst predictor for our task as presented in Figure 13.

Females predictive models performance is reported in Table V. Also Random Forest algorithm had the best results when trained on females trait specific feature resulting in a higher accuracy in predicting Openness, Extraversion, Agreeableness and Neuroticis from their facial features. Support Vector Regressor was the best performing algorithm in predicting Conscientiousness personality trait with 0.07 accumulated mean squared error as presented in Figure 12. In general, SVR and Random Forest were the best performing algorithms to predict the personality traits for both genders. On the females dataset, Bagging were the worst prediction algorithm.

TABLE IV. PREDICTIVE PERFORMANCE OBSERVED USING MEAN SQUARED ERROR USING SVR, RANDOM FOREST, BAGGING AND ADABOOST FOR MALES TWITTER SAMPLE.

Trait	SVR	R_Forest	Bagging	AdaBoost
Openness	0.10619	**0.09882**	0.3265	0.2983
Conscientiousness	0.07426	**0.06649**	0.2621	0.2564
Extraversion	0.06683	**0.05485**	0.2530	0.2482
Agreeableness	0.08650	**0.08444**	0.2721	0.2459
Neuroticism	**0.07694**	0.09254	0.2029	0.1974

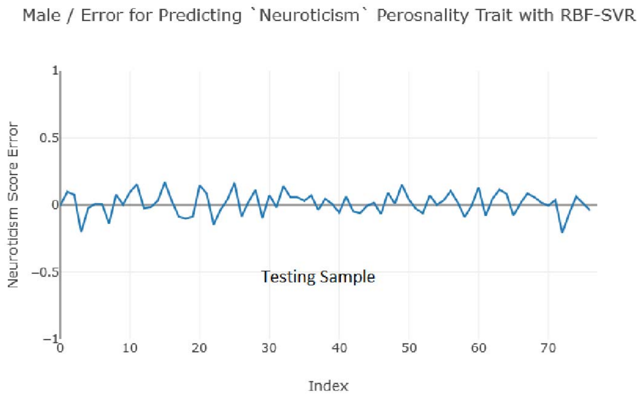

Fig. 11. Mean Squared Error Results for Neuroticism Personality Trait With Support Vector Regressor among RBF Kernel at the Males Samples

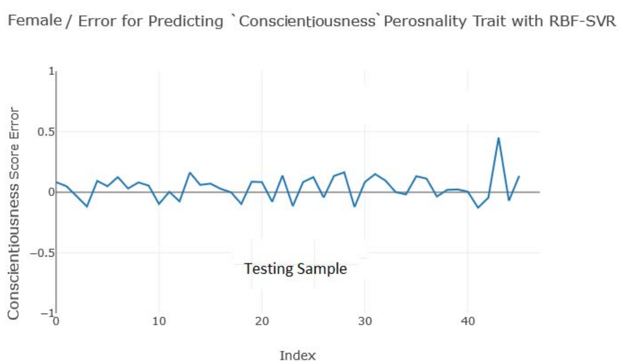

Fig. 12. Mean Squared Error Results for Conscientiousness Personality Trait With Support Vector Regressor among RBF Kernel at the Females Samples.

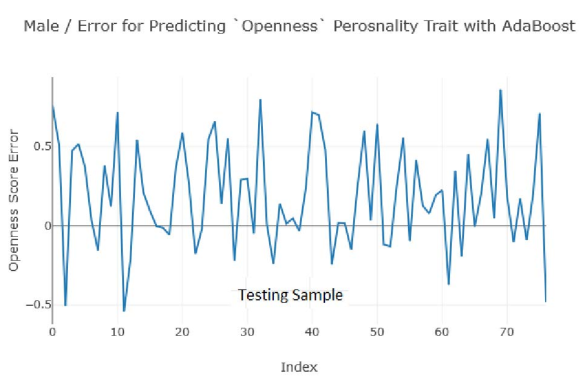

Fig. 13. Mean Squared Error Results for Openness to Experience Personality Trait With Adaptive Boosting Algorithm at Males Samples.

TABLE V. PREDICTIVE PERFORMANCE OBSERVED USING MEAN SQUARED ERROR USING SVR, RANDOM FOREST, BAGGING AND ADABOOST FOR FEMALES TWITTER SAMPLE

Trait	SVR	R_Forest	Bagging	AdaBoost
Openness	0.1170	**0.08266**	0.5307	0.4773
Conscientiousness	**0.0751**	0.07991	0.4312	0.4067
Extraversion	0.0706	**0.06714**	0.4611	0.4172
Agreeableness	0.1000	**0.08966**	0.4267	0.4054
Neuroticism	0.1029	**0.09384**	0.3230	0.3117

V. FUTURE WORK

Our research is only using one single profile image from users to make the final prediction, future work efforts would be a promising to conduct and investigate study with experience sampling where uploaded images over time from users are collected and analyzed. Further, more enhanced and engineered generated features from users profile pictures will be investigated once we manage to increase our initial training samples. Although the information revealed within the profile picture is only a small part of the total information a user leaves on social platforms, the generated models from this study can be integrated into a learning ensemble that consider other relevant information such as Likes and Posts into further and more sophisticated prediction framework.

VI. CONCLUSION

Considering the fact that a significant effort on cross-domain research in data science and psychology has been achieved to grasp the human online characteristics, estimating individuals personality scores from their face and appearance is still mostly not fully explored and this research effort contributes to this direction and the goal is to investigate the best possible features to build automatic personality estimation models. Our research experiments are made while ethical and privacy issues are concerned, the goal is to raise the awareness between social media users of what third parties can reveal about their private traits from what they share and behave in various social networking platforms.

The final models utilizes two distinct approaches to select feature sets and evaluates four different types of machine learning algorithms. The final models are able to accurately estimate users personality scores by analyzing a huge set of combination among facial features with state-of-the-art machine learning models. The evaluation of our best Random Fores models results reveals that there is a considerable connection among users personality and the photo they choose as profile pictures. We concluded that human gender has an immense role in building personalized personality prediction models.

APPENDIX

TABLE VI: Feature Sets for Males and Females Selected by Pearson Correlation Coefficient and Gradient Boosting Feature Importances. In total, there is 50 unique Facial Feature returned by *Face++*.

Set	Type	Features' Name
HPP_A	Numerical	*A Headpose_Pitch_Angle*
Mb_V	Numerical	*Motionblur_Value*
Face_Q	Numerical	*Face_Quality*
GB_V	Numerical	*Gaussianblur_Value*
BM_S	Numerical	*Beauty_Male_Score*
RES_O	Numerical	*Right_Eye_Status_Occlusion*
FR_W	Numerical	*Face_Rectangle_Width*
LESNGE_O	Numerical	*Left_Eye_Status_Normal_Glass_Eye_Open*
RESNGE_C	Numerical	*Eight_Eye_Status_Normal_Glass_Eye_Close*
FR_L	Numerical	*Face_Rectangle_Left*
FR_H	Numerical	*Face_Rectangle_Height*
FR_W	Numerical	*Face_Rectangle_Width*
LESNGE_C	Numerical	*Left_Eye_Status_Normal_Glass_Eye_Close*
LEGV_C	Numerical	*Left_Eye_Gaze_Vector_Y_Component*
REGVY_C	Numerical	*Right_Eye_Gaze_Vector_y_Component*
RESNGE_O	Numerical	*Right_Eye_Status_Normal_Glass_Eye_Open*
SS_S	Numerical	*SkinStatus_Stain*
REGVZ_C	Numerical	*Right_Eye_Gaze_Vector_Z_Component*
SSD_C	Numerical	*SkinStatus_Dark_Circle*
SS_H	Numerical	*SkinStatus_Health*
LESD_G	Numerical	*Left_Eye_Status_Dark_Glasses*
REGPX_C	Numerical	*Right_Eye_Gaze_Position_X_Coordinate*
MSO_O	Numerical	*MouthStatus_Other_Occlusion*
REGPY_C	Numerical	*Right_Eye_Gaze_Position_Y_Coordinate*
LESNGE_C	Numerical	*Left_Eye_Status_No_Glass_Eye_Close*
RESNoGE_C	Numerical	*Right_Eye_Status_No_Glass_Eye_Close*
E_Surprise	Numerical	*Emotion_Surprise*
REGVX_C	Numerical	*Right_Eye_Gaze_Vector_X_Component*
HPY_A	Numerical	*Headpose_Yaw_Angle*
MS_O	Numerical	*MouthStatus_Open*
LEGVX_C	Numerical	*Left_Eye_Gaze_Vector_X_Component*
BF_S	Numerical	*Beauty_Female_Score*
E_Surprise	Numerical	*Emotion_Surprise*
E_Neutral	Numerical	*Emotion_Neutral*
E_Anger	Numerical	*Emotion_Anger*
E_Happiness	Numerical	*Emotion_Happiness*
E_Fear	Numerical	*Emotion_Fear*
E_Disgust	Numerical	*Emotion_Disgust*
Age_value	Numerical	*Age_Value*
MS_C	Numerical	*MouthStatus_Close*
MSSM_R	Numerical	*MouthStatus_Surgical_Mask_or_Respirator*
LEGVZ_C	Numerical	*Left_Eye_Gaze_Vector_Z_Component*
LEGPY_C	Numerical	*Left_Eye_Gaze_Position_y_coordinate*
FQuality_V	Numerical	*FaceQuality_Value*
LESNGE_O	Numerical	*Left_Eye_Status_No_Glass_Eye_Open*
SS_A	Numerical	*SkinStatus_Acne*
LES_O	Numerical	*Left_Eye_Status_Occlusion*
RESNGE_O	Numerical	*Right_Eye_Status_No_Glass_Eye_Open*
FR_T	Numerical	*Face_Rectangle_Top*
RESNGE_C	Numerical	*Right_Eye_Status_Normal_Glass_Eye_Close*

REFERENCES

[1] We are Social and Hootsuite, "Digital in 2017 Global Overview", In slides, 2018-7-7, https://www.slideshare.net/wearesocialsg/digital-in-2017-global-overview.

[2] R. Hu, P. Pu, "Enhancing collaborative filtering systems with personality information." In Proceedings of the fifth ACM conference on Recommender systems (RecSys '11). ACM, New York, NY, USA, 197-204. DOI=http://dx.doi.org/10.1145/2043932.2043969.

[3] M. Sandra C., M. Kosinski, G. Nave, D. Stillwell, "Psychological targeting as an effective approach to digital mass persuasion." Proceedings of the national academy of sciences 114, no. 48 (2017): 12714-12719.

[4] R. Bin Tareaf, P. Berger, P. Hennig, C. Meinel, "Personality Exploration System for Online Social Networks: Facebook Brands As a Use Case." In 2018 IEEE/WIC/ACM International Conference on Web Intelligence (WI), pp. 301-309. IEEE, 2018.

[5] C. Li, W. Wu, L. He, "How personality influences users' needs for recommendation diversity?." In CHI'13 Extended Abstracts on Human Factors in Computing Systems, pp. 829-834. ACM, 2013.

[6] H. Sajanee,J. Kientz, "Personality and persuasive technology: an exploratory study on health-promoting mobile applications." In International conference on persuasive technology, pp. 150-161. Springer, Berlin, Heidelberg, 2010.

[7] F. Bruce, M. Schedl, "Personality-based user modeling for music recommender systems." In Joint European Conference on Machine Learning and Knowledge Discovery in Databases, pp. 254-257. Springer, Cham, 2016.

[8] F. Bruce, M. Graus, A. Vall, M. Tkalcic, M. Schedl, "The Influence of Users' Personality Traits on Satisfaction and Attractiveness of Diversified Recommendation Lists." In EMPIRE@ RecSys, pp. 43-47. 2016.

[9] L. Michael, B. Ferwerda, "Personalizing online educational tools." In Proceedings of the 2017 ACM Workshop on Theory-Informed User Modeling for Tailoring and Personalizing Interfaces, pp. 27-30. ACM, 2017.

[10] R. Bin Tareaf, P. Berger, P. Hennig, J. Jung, C. Meinel, "Identifying audience attributes: predicting age, gender and personality for enhanced article writing." In Proceedings of the 2017 International Conference on Cloud and Big Data Computing, pp. 79-88. ACM, 2017.

[11] P. Costa, R. McCrae, "The revised neo personality inventory (neo-pi-r)." The SAGE handbook of personality theory and assessment 2.2, pp. 179-198, 2008.

[12] M. Kosinski, D. Stillwell, T. Graepel, "Private traits and attributes are predictable from digital records of human behavior." In Proceedings of the National Academy of Sciences, Vol. 110, 2013

[13] Q. Daniele, M. Kosinski, D. Stillwell, J. Crowcroft. "Our twitter profiles, our selves: Predicting personality with twitter." In 2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing, pp. 180-185. IEEE, 2011.

[14] R. Bin Tareaf, B. Philipp, H.Patrick, M. Christoph, "ASEDS: Towards Automatic Social Emotion Detection System Using Facebook Reactions." In 2018 IEEE 20th International Conference on High Performance Computing and Communications; IEEE 16th International Conference on Smart City; IEEE 4th International Conference on Data Science and Systems (HPCC/SmartCity/DSS), pp. 860-866. IEEE, 2018.

[15] R. Bin Tareaf, B. Philipp, H.Patrick, M. Christoph, "Towards Automatic Personality Prediction Using Facebook Like Categories." arXiv preprint arXiv:1812.04346 (2018).

[16] L. Leqi, D. Preotiuc-Pietro, Z. Samani, M. Moghaddam, L. Ungar, "Analyzing personality through social media profile picture choice." In Tenth international AAAI conference on web and social media. 2016.

[17] F. Bruce, M. Schedl, M. Tkalcic, "Predicting personality traits with instagram pictures." In Proceedings of the 3rd Workshop on Emotions and Personality in Personalized Systems 2015, pp. 7-10. ACM, 2015.

[18] Megvii, I. "Face++ research toolkit." (2013).

[19] Richter, F. (2017). Snapchat crowned number 1 by American teens. Statista. Retrieved from https://www.statista.com/chart/4823/teenagers-favorite-social-networks

[20] S. Marcin, M. Tkali, B. Ferwerda, M. Schedl. "Fusing social media cues: personality prediction from twitter and instagram." In Proceedings of the 25th international conference companion on world wide web, pp. 107-108. International World Wide Web Conferences Steering Committee, 2016.

[21] N. Jie, L. Huang, P. Cui, Z. Li, Y. Yan, Z. Wei, W. Zhu. "Social media profiler: Inferring your social media personality from visual attributes in portrait." In Pacific Rim Conference on Multimedia, pp. 640-649. Springer, Cham, 2016.

[22] C. Marco, A. Vinciarelli, C. Segalin, A. Perina. "Unveiling the multimedia unconscious: Implicit cognitive processes and multimedia content analysis." In Proceedings of the 21st ACM international conference on Multimedia, pp. 213-222. ACM, 2013.

[23] D. Stillwell, M. Kosinski. "myPersonality Project website", 2016-11-25, http://mypersonality.org/.

[24] P. Booth. "Francis (2007) Pennebaker, JW, Booth, RJ, & Francis, ME (2007)." LIWC 2007: Linguistic Inquiry and Word Count.

[25] Megvii, I. "Face++ research toolkit." (2013).

[26] T. Hastie, R. Tibshirani, J. Friedman, "Unsupervised learning." The elements of statistical learning. In Springer, New York, NY, pp.485-585, 2009.

[27] J. Vert, K. Tsuda, B. Schlkopf,"A Primer on Kernel Methods", 2004-07-27, In Url http://members.cbio.mines-paristech.fr/ jvert/publi/04kmcbbook/kernelprimer.pdf.

[28] M. Gregory , S. Finn, L. Eyde, G. Kay, K. Moreland, R. Dies, E. Eisman, T.Kubiszyn, Ge. Reed. "Psychological testing and psychological assessment: A review of evidence and issues." American psychologist 56, no. 2 (2001): 128.