SEMINAR

1. **Using textual data for Personality Prediction:A Machine Learning Approach**

<https://ieeexplore.ieee.org/abstract/document/9036220>

Ne mogu pristupiti iz nekog razloga.

KEYWORDS:

SUMMARIZATION:

# Personality Prediction System Based on Signatures Using Machine Learning

<https://iopscience.iop.org/article/10.1088/1757-899X/879/1/012068/pdf>

KEYWORDS: signatures, graphology, Principle Component Analysis, Support Vector Machine, image, confusion matrix, presonality prediction, machine learning

SUMMARIZATION:

Predicting presonality based on signature patterns using machine learning. Model training and testing. Four features: curve start, end streak, middle stroke, underline. **Principle Component Analysis and Support Vector Machine.** Accuracy value is 71%, good accuracy. Used for employee recruitment. Consists of 10 classes, each one represents a personality. The input is a digital image. 530 signatures, 53 respondents who write 10 signatures. Image training and testing, image processing, vertical and horizontal segmentation, feature extraction, SVM classification. Confusion matrix method.

1. Personality Prediction System from Facebook Users

<https://reader.elsevier.com/reader/sd/pii/S1877050917320537?token=54186579D8DA5C959BEFDCDAB6C229209660AE33F6ADE59ADD7FA6DB152631A1DD197155E71294CEEE91D365ADFABF44&originRegion=eu-west-1&originCreation=20221219150952>

KEYWORDS: personality prediction, Facebook, deep learning, Big Five Model Personality

SUMMARIZATION:

Using the Big Five Model Personality. Deep learning is used instead of machine learning. Outperforms previous similar researches with the average accuracy of 74.17%. Two datasets, prediction system used features like LIWC and SPLICE. SNA – Social Network Analysis. The second approach used open vocabulary approach. Deep learning was not used for this purpose before. Deep learning can improve the accuracy even if the accuracy is still quite low for some traits – possibly due to small number of datasets used in this study. For future studies - XGBoost algorithm.

1. **Survey Analysis of Machine Learning Methods for Natural Language Processing for MBTI Personality Type Prediction**

<http://cs229.stanford.edu/proj2017/final-reports/5242471.pdf>

KEYWORDS: survey, machine learning, natural language processing, MBTI, social media, Python, Stohastic Gradient Descent, Softmax classifier, Naïve Bayes, Support Vector Machines, Deep Learning

SUMMARIZATION:

Classifying someones MBTI personlity type based on their social media posts, using NLP with machine learning techniques. Variety of methods, classical Supervised Learning and deep learning. Python and libraries. Stohastic Gradient Descent, Softmax classifier used for baseline, outperforms randomly choosing classes. Naïve Bayes for text classification. Support Vector Machines (SVMs) are very wellperforming, robust, and customizable supervised learning algortihms, L2-regularization for better generalization. Regularized SVM on individual personality types yields better accuracy than Baseline and Naïve Bayes models, Deep Learning outperforms SVM. Overall final accuracy doesn’t surpass 50 percent. Moving forwards, incorporate richer data and features for stronger understanding of the input text and improved performance. Try unsupervised learning and other mechnisms of deep learning.

# Machine prediction of personality from Facebook profiles

<https://ieeexplore.ieee.org/abstract/document/6302998/authors#authors>

Ne mogu ni do ovog.

# Predicting Dark Triad Personality Traits from Twitter Usage and a Linguistic Analysis of Tweets

<https://ieeexplore.ieee.org/abstract/document/6406767>

Ni do ovog.

# Personality Traits Identification Using Rough Sets Based Machine Learning

<https://ieeexplore.ieee.org/abstract/document/6724349>

KEYWORDS:

SUMMARIZATION:

1. **Cross-Cultural Personality Prediction based on Twitter Data**

<http://www.jsoftware.us/vol12/300-TE040.pdf>

KEYWORDS: cross-cultural, personality prediction, Twitter, regression analysis, social media mining, suprevised learning

SUMMARIZATION:

Cross-cultural personality prediction – a new research topic. South Korea and Turkey. IBM Presonality Insight. Most studies apply the Big Five Model. Based on social media data. Developing software systems for collecting Twitter data. Machine learning algorithms for 5 dimensions of personality. After the best models are identified for each culture, cross-cultural prediction expreiments will be performed.

1. **Improving Intelligent Personality Prediction using Myers-Briggs Type Indicator and Random Forest Classifier**

<https://pdfs.semanticscholar.org/11f9/ccf33ec4afd03678338c1de7c296d5693740.pdf>

KEYWORDS: personality prediction, MBTI, Random Forest Classifier, Twitter

SUMMARIZATION:

Machine Learning technique usingg Random Forest Classifier to automatically predict peoples personality based on MBTI. Random Forest Classifier performs better than different tree machine learning algorithms. MBTI is more robust than others in terms of validity and reliability. Using Twitter. Random Forest Classifier advantages: robust,highly accurate,no overfitting, classification and regression,… disadvantages: slow prediction generation process, challenging to interpret,… Explorazory data analysis. Random Forest model has better accuracy (100%) in all four dimensions of MBTI personality types compared to other machine learning models. Try other social media platforms. Try XGBoost algorithm and deep learning.

# Predicting personality from patterns of behavior collected with smartphones

<https://www.pnas.org/doi/epdf/10.1073/pnas.1920484117>

KEYWORDS: smartphones, personality, behavior, Big Five personality dimensions, machine learning, digital footprints, mobile sensing, privacy

SUMMARIZATION:

Predicting the extent to which individuals Big Five personality dimensions can be predicted on the basis of six different classes of behavioral information collected via sensor and log data harvested from smartphones. Benefits and dangers presented by collection and modeling of behavioral data obtained from smartphones. Specific classes of behavior like app usage, music consumption, social behavior were distinctively informative about the different Big Five trait dimensions. Models were able to predict personality on the broad domain level and the narrow facet level for oppeness, conscientiousness and extraversion. Agreeableness could not be predicted. Communication and social behavior were the most important. Greater prediction with more sensors.

# User personality prediction based on topic preference and sentiment analysis using LSTM model

<https://www.sciencedirect.com/science/article/abs/pii/S0167865520302919>

KEYWORDS: personality,topic,sentiment analysis,LSTM model, deep learning, Facebook

SUMMARIZATION:

Long Short-Term Memory model to predict the personality characteristics of social network uses. Attention-based LSTM model can achieve better results than the currently popular methiods in the recognition of user personality traits, good generalization ability. o explore the effective methods of user personality prediction, this paper takes the Big Five model theory and LDA theme model as a priori knowledge, introduces an attention mechanism and builds an Attention-based LSTM model that combines thematic and emotional features. Based on the original text information, this model converts the users' theme preferences and text sentiment features into attention information. It is combined with the LSTM model in different forms to predict the personality characteristics of social network users. Since the data set used in this paper is a Facebook corpus containing user personality tags, this model is a supervised deep learning model. During the construction of the model, the original text information needs to be input into the LSTM model. The thesis uses real Facebook data to construct the model.

# Personality Prediction of Social Network Users

<https://ieeexplore.ieee.org/abstract/document/8253041>

ZADATAK

• Što je Personality Prediction i zašto se koristi?

• Koji su glavni ciljevi upotrebe strojnog učenja u Personality Predictionu i kako ono poboljšava predviđanje osobnosti?

• Koji su glavni izazovi s kojima se suočavaju istraživači prilikom korištenja strojnog učenja za Personality Prediction i kako se oni prevazilaze?

• Koji su glavni algoritmi strojnog učenja koji se koriste za Personality Prediction i koje su njihove prednosti i nedostaci?

• Kako se koriste podaci u strojnom učenju za Personality Prediction i koje su najčešće korištene metode obrade podataka?

• Koji su glavni rezultati istraživanja o upotrebi strojnog učenja za Personality Prediction i koji su najvažniji zaključci koje se iz njih mogu izvući?

• Koje su buduće perspektive i mogućnosti upotrebe strojnog učenja za Personality Prediction i koje su glavni izazovi s kojima će se suočiti istraživači u budućnosti?

**Using textual data for Personality Prediction:A Machine Learning Approach**

**•** **Što je Personality Prediction i zašto se koristi?**

Predicting personality with the help of data through social

media is a promising approach as this method does not require

any questionnaires to be filled by users thus reducing time and

increasing credibility.

Prediction of personality is an area of study where

person gets categorized in a class according to his/her

personality.

**• Koji su glavni ciljevi upotrebe strojnog učenja u Personality Predictionu i kako ono poboljšava predviđanje osobnosti?**

* Predicting personality with the help of data through social

media is a promising approach as this method does not require

any questionnaires to be filled by users thus reducing time and

increasing credibility.

* All these traditional

methods of personality prediction use questionnaire for

personality prediction. Filling a lengthy questionnaire is time

consuming and tedious job.

* Advanced machine learning algorithm - AdaBoost has

been used in [8] for prediction. Boosting algorithms are used

to gain highest accuracy and high performance.

**• Koji su glavni izazovi s kojima se suočavaju istraživači prilikom korištenja strojnog učenja za Personality Prediction i kako se oni prevazilaze?**

- dataset creation and tools that helpo with that

- social media trolling and detection processes

- data preprocessing

**• Koji su glavni algoritmi strojnog učenja koji se koriste za Personality Prediction i koje su njihove prednosti i nedostaci?**

* Advanced machine learning algorithm - AdaBoost has

been used in [8] for prediction. Boosting algorithms are used

to gain highest accuracy and high performance. Hence we

have employed AdaBoost with other algorithms like LDA

and Multinomial Naïve Bayes to analyze accuracies of

algorithms

* Linear regression and support vector regression -> linear regression better for prediction
* Random forest, simple logistics, J48
* Binary SVM for mental disorder detection
* CNN and DCNN
* Tree, SVM, Ensamble, kNN and ANN for ECG authentication and gender recognition
* NLP processing
* Approaches used to execute this include LIWC

software, Hogan Development Survey, Ordinary Least

Square regression and Least Absolute Shrinkage and

Selection Operator regression.

* Thus from all the graphical representations it is clear that

Multinomial Naïve Bayes outperforms LDA and AdaBoost

in terms of accuracy, precision, recall and F1-score.

**• Kako se koriste podaci u strojnom učenju za Personality Prediction i koje su najčešće korištene metode obrade podataka?**

* It is easier to gather a dataset from social media like Twitter or Facebook and use it for prediction.
* Dataset for this is generated using Amazon Mechanical Turk.
* NLP processing such as removing stop words, tokenization, stemming is applied to the extracted posts. Thus negative and positive sentiments according to word usage are determined
* Trolling detection
* Csv storing and preprocessing
* Python library Tweepy, for processing tweets
* Tokenization, stemming and removing stop words

**• Koji su glavni rezultati istraživanja o upotrebi strojnog učenja za Personality Prediction i koji su najvažniji zaključci koje se iz njih mogu izvući?**

* Hence personality prediction helps in such scenarios

where it is easy to compute sentiment of a tweet or post

posted by user by applying algorithms.

As discussed in our work we have applied Multinomial

Naïve Bayes, AdaBoost and LDA to compare which

algorithm has higher relevance. Thus, according to our

results it is found that Multinomial Naïve Bayes has highest

accuracy of 73.43, precision of 0.7, and recall of 0.71 and

F1-score of 0.72. Future scope aims in improving accuracy

of algorithm.

**• Koje su buduće perspektive i mogućnosti upotrebe strojnog učenja za Personality Prediction i koje su glavni izazovi s kojima će se suočiti istraživači u budućnosti?**

* There is a scope of research in combining

multiple tests of personality to find most accurate class

labels. Also more focus can be thrown on real-time data

which can have significance with real world. Combining

Machine Learning algorithms can be useful in improving

accuracy. Lastly, there is a scope to work in multimodal

approach of prediction in which different biomedical signals

can be considered.

* Thus, according to our

results it is found that Multinomial Naïve Bayes has highest

accuracy of 73.43, precision of 0.7, and recall of 0.71 and

F1-score of 0.72. Future scope aims in improving accuracy

of algorithm.

**Machine Prediction of Personality from Facebook Profiles**

• **Što je Personality Prediction i zašto se koristi?**

* Predicting personality traits, Big Five personality model
* Rank individuals in terms of the Big Five
* Advertisment
* Improving members social capital
* The significance of these results is that when considering large

groups of users, automatic analysis can be used to identify those

with specific personality traits. This can be scaled up far more easily

than manual human assessment, and can be used for targeting specific

groups of users for advertising, social engineering attacks, or finding influential users.

**• Koji su glavni ciljevi upotrebe strojnog učenja u Personality Predictionu i kako ono poboljšava predviđanje osobnosti?**

* In this study, we apply datamining and machine learning techniques to predict users’ personality traits (specifically, the traits of the Big Five personality model) using only demographic and text-based attributes extracted from their profiles.
* Using raw statistical tools, some useful patterns (such as a connection between number of friends and extroversion) were discovered.
* These studies examine such questions as how user demographics and knowledge pertain to online behavior [1], [7], how users cope with privacy challenges [3], and what specific threats to privacy most concern users

**• Koji su glavni izazovi s kojima se suočavaju istraživači prilikom korištenja strojnog učenja za Personality Prediction i kako se oni prevazilaze?**

* Privacy concerns
* Most concerns pertaining to social networking posts focus on raw

demographic information or specific damaging posts. For example,

religious or political affiliation may impact on how an employer

views a potential employee, as might an inflammatory post or

salacious photograph. However, relatively little attention has focused

on aggregating information from posts and profile text to create a

broad picture of the user.

* In many application domains, too much data is generated to be

easily processed by humans. Although patterns exist, they require

connecting vast quantities of data together in ways which would

take years to be discovered manually. The field of data mining

encompasses a wide range of techniques for extracting information

from large datasets [22].

**• Koji su glavni algoritmi strojnog učenja koji se koriste za Personality Prediction i koje su njihove prednosti i nedostaci?**

* A number of different numeric

prediction models have been employed throughout the literature,

including linear regression, REPTree, and decision tables. These

three models were chosen for use in this study because preliminary

analysis showed they had the greatest performance. All models built

in this study were implemented using the WEKA machine learning

toolkit [8], using default values for all parameters except as noted

below.

* Linear regression – simple numeric prediction
* REPTTree
* Decision tables
* Overall, using the REPTree-based

models, we found that for every combination of personality trait and

direction (top or bottom), we were able to predict the top 10% with

34.5% accuracy, exceeding 21.8%, the accuracy of simply ordering

the users based on the best-performing independent attribute.

**• Kako se koriste podaci u strojnom učenju za Personality Prediction i koje su najčešće korištene metode obrade podataka?**

* In this study, we apply datamining

and machine learning techniques to predict users’ personality

traits (specifically, the traits of the Big Five personality model) using

only demographic and text-based attributes extracted from their profiles.

* Classification, clustering

Rather than directly turning different aspects

of the profiles into categorical or numeric data, these attributes arose

from taking the users’ profiles, photo descriptions, and public posts

and processing them using the Linguistic Inquiry and Word Count

software package [16]. This tool analyzes text to pull out “language

dimensions” including positive and negative emotional content, selfreference,

key topics, writing style, and more.

**• Koji su glavni rezultati istraživanja o upotrebi strojnog učenja za Personality Prediction i koji su najvažniji zaključci koje se iz njih mogu izvući?**

* Our results show that when using certain models, we

can find the top 10% most Open individuals with nearly 75% accuracy,

and across all traits and directions, we can predict the top 10% with

at least 34.5% accuracy (exceeding 21.8%, which is the best accuracy

when using just the best-performing profile attribute). These results have

privacy implications in terms of allowing advertisers and other groups to

focus on a specific subset of individuals based on their personality traits.

* The significance of these results is that when considering large

groups of users, automatic analysis can be used to identify those

with specific personality traits. This can be scaled up far more easily

than manual human assessment, and can be used for targeting specific

groups of users for advertising, social engineering attacks, or finding

influential users. While many are aware of the risks associated with

posting inappropriate photographs or posts on social networking sites,

this research demonstrates that even innocuous-seeming bibliographic

and status information can be used to discern information which users

may prefer to remain hidden.

**• Koje su buduće perspektive i mogućnosti upotrebe strojnog učenja za Personality Prediction i koje su glavni izazovi s kojima će se suočiti istraživači u budućnosti?**

* Future research will explore additional techniques to interpret

users’ personalities through automated data mining analysis of Facebook

profiles (for example, by applying feature selection to reduce

the number of independent attributes), as well as examine additional

datasets from other social networks.

* In this preliminary study, all 111 attributes (31 demographic + 80

text-based) were used to build models. Future work may consider the

use of feature selection to reduce the size of the feature space.

**Predicting Dark Triad Personality Traits from Twitter Usage and a Linguistic Analysis of Tweets**

**• Što je Personality Prediction i zašto se koristi?**

* To date, studies have typically focused on the Big Five traits of personality, but one area which is relatively unexplored is that of the anti-social traits of narcissism, Machiavellianism and psychopathy, commonly referred to as the Dark Triad. This study explored the extent to which it is possible to determine anti-social personality traits based on Twitter use. This was performed by comparing the Dark Triad and Big Five personality traits of 2,927 Twitter users with their profile attributes and use of language.
* While predictive models may be unsuitable for predicting an individual’s personality, they may still be of practical importance when models are applied to large groups of people, such as gaining the ability to see whether anti-social traits are increasing or decreasing over a population.

**• Koji su glavni ciljevi upotrebe strojnog učenja u Personality Predictionu i kako ono poboljšava predviđanje osobnosti?**

* Through the use of crowd sourced machine learning algorithms, we show that machine learning provides useful prediction rates, but is imperfect in predicting an individual’s Dark Triad traits from Twitter activity. While predictive models may be unsuitable for predicting an individual’s personality, they may still be of practical importance when models are applied to large groups of people, such as gaining the ability to see whether anti-social traits are increasing or decreasing over a population.

**• Koji su glavni izazovi s kojima se suočavaju istraživači prilikom korištenja strojnog učenja za Personality Prediction i kako se oni prevazilaze?**

**• Koji su glavni algoritmi strojnog učenja koji se koriste za Personality Prediction i koje su njihove prednosti i nedostaci?**

**• Kako se koriste podaci u strojnom učenju za Personality Prediction i koje su najčešće korištene metode obrade podataka?**

**• Koji su glavni rezultati istraživanja o upotrebi strojnog učenja za Personality Prediction i koji su najvažniji zaključci koje se iz njih mogu izvući?**

**• Koje su buduće perspektive i mogućnosti upotrebe strojnog učenja za Personality Prediction i koje su glavni izazovi s kojima će se suočiti istraživači u budućnosti?**