TECHNIUM
Technium Vol. 23, pp.6-17 (2024)
ISSN: 2668-778X
www.techniumscience.com

Seamless Transition: Mapping Smartphone-Derived Personality Profiles to Business Intelligence

Rzeźnik Jakub[a,b,1], Krzemińska Izabella[a,2]

[a] Orange Poland, Al. Jerozolimske 160, , Warszawa, 02-691, Poland
[b] Warsaw School of Economics, Al. Niepodległości 162, , Warszawa, 02-691, Poland

Abstract. This study explores the feasibility of transferring a personality prediction model, trained on smartphone data, to a Business Intelligence (BI) system utilizing call logs. By leveraging the Big Five personality traits model and machine learning techniques, the research addresses the challenge of harmonizing disparate data sources—rich data about smartphone applications usage and system settings, and aggregated call log data from a BI warehouse. Through meticulous data preprocessing and feature engineering, the study evaluates the performance of various machine learning algorithms, demonstrating that traits like Extraversion and Conscientiousness can be reliably predicted. The study also introduces innovative methods for verifying predictions on unlabeled BI data, including consistency over time and similarity of distributions. The findings underscore the potential of integrating personality insights into BI systems, also highlighting the ethical considerations and data privacy challenges inherent in such applications. This research contribution lays the groundwork for future advancements in personality prediction models and their practical business applications.

Keywords: user experience, mobile application, personalization of service, personality derived from data, machine learning

1. Introduction and Motivation

Over the last decade, researchers have been studying methods of determining a user's personality from their digital traces, and this research tendency began with the expansion of social media (Kosinski et al., 2015). As a result, personality is often inferred from social media data, primarily using large text datasets available as open data on platforms such as Twitter and MyPersonality (Bin Tareaf et al., 2019; Khan et al., 2020). Some researchers have also attempted to predict personality based on profile pictures on social media (Liu et al., 2016). There have been attempts to explore other types of data, such as logs from phone calls (Montjoye et al., 2013), application events (Xu et al., 2016), eye movements (Berkovsky et al., 2019), and wearable events (Kalimeri et al., 2013). However, some studies based on social media data are impossible to repeat due to legal restrictions on access to personal data, which limits their use outside the service (e.g. Facebook, Twitter). Despite this, the research is still valuable from a cognitive perspective.

Most studies are based on a large amount of data from a service's history. However, it appears that some researchers focus more on improving machine learning methods than on the usefulness of the models created in practice, for example (Bin Tareaf et al., 2019; Khan et al., 2020). Using such models for business purposes is also limited to commercial personalization in social media or service recommendation (Ning et al., 2019; Matz et al., 2017).

The Big Five Factors model, investigated and described by Costa Jr and McCrae (1992) alone is the most commonly used model for personality in the context of digital data, as it was researched by (Krzeminska, 2022). Its popularity can be evidenced by the existence of the Cybernetic Big Five Factors model developed by (DeYoung, 2015).

What is important due to the proposed approach is the fact that in the analyzed studies in the literature, the personality profile is usually a set of discrete binary variables. Although it is convenient in modelling (the problem of unbalanced classes disappears), it is not justified in the case of personalization. Here, features are to be used to identify those users who significantly differ from a typical user in terms of needs and motivations to act. And the use of binary classes is less likely to reflect variation in human behaviour (Stajner and Yenikent, 2021; Krzeminska, 2022).

The main reason why the business is interested in the deep understanding of the customer on very individual level is the basic feature of a modern agile enterprise: the flexible adaptation of final products to the current needs of customers.

* This document is the results of the research project funded by Orange Innovation Poland, CURIE Research Program lead by Michal Szczerbak in Pawel Tuszynski, AI Competence Center Team

Jakub.Rzeznik@orange.com (R. Jakub); Izabella.Krzeminska@oramge.com (K. Izabella)
ORCID(s): 0000-0003-0078-7395 (R. Jakub); 0000-0002-6913-7393 (K. Izabella)

In an environment of high volatility, strategies have been developed to counteract problems resulting from the fact that it takes too long to deliver the product to the customer after creating a product concept. The literature on the subject describes various logistics strategies of enterprises, and one of them is *postponing* - late personalization or delayed product differentiation. The main reason for the delay strategy is to gain a competitive advantage by providing the customer with the most up-to-date personalized product (Swaminathan, 2003).

According to Brdulak (2014), one of the strategies for managing uncertainty is to create flexible and self-regulating systems based on artificial intelligence. Technologies enable the development of such a self-adaptable system using available data to create knowledge about the customer. Such a system also includes the possibility of postponing the creation of the final shape of the product until the service is installed at the customer's device.

Two years ago, a base model was created to predict a user's personality profile from smartphone data Krzeminska (2022). This model was unique because it only required a single feature enumeration to be calculated and did not need to track the user over time. The idea was attractive from a business perspective because it allowed beneficial knowledge about the user to be obtained during onboarding to the mobile application. The model was calculated locally and did not send any data outside the phone. The entire model creation process, data, and quality parameters are described in previous publications by Krzeminska.

When designing user tests to prove the concept for a telecommunications operator, a question arose whether it is possible to calculate a user's profile from their call logs history. This approach would make it possible to gain knowledge about the user without requiring any other category of data. (Krzeminska, 2022; Krzeminska and Szmydt, 2021).

When designing user tests to demonstrate proof of concept for a telecommunications operator, a question arose about whether it was possible to calculate a user's profile from their connection history. This procedure would enable adapting the customer care model to suit individual users better and to provide more opportunities to verify diverse behaviors. Having such a profile prediction in the BI database would also open up new opportunities to create channels and service scenarios consistent with users' needs.

One of our research questions was whether the classification model trained on cell phone data could be transferred to a BI database where only similar information was available to train a new model. The main concern was whether the models trained on a targeted sample of approximately 6,000 people would replicate well in a database of several million customers. It should be noted that the model calculated on the smartphone was based on different categories of data, including applications, system information, phone usage settings, and usage characteristics. In contrast, only the user's call history information was available in the BI database.

Since the business-psychological concept remained the same as in our previous research (see: Krzeminska (2022)), this article presents a limited literature review mainly concerning model transfer methods. It also provides a detailed description of the model creation process carried out on new and different data, including verifying data adequacy. Furthermore, it presents the obtained model quality metrics for some personality traits, explaining why it was not possible for two of them. Finally, the article discusses experiments, conclusions, and recommendations for possible paths of continuation of research.

2. Research Objectives

The above motivation defines the main research problem analyzed in this publication - transfer of a predictive model trained on a limited user base to a database of different origin with a several million users. To address this issue, the following research questions need to be answered: RQ1. Can profiles be predicted using models trained on data from sources other than the data used for prediction? And RQ2. How to measure the quality of such a forecast?

3. Literature Review

Integrating mobile data predictive models with business intelligence (BI) systems, especially for personality profiling based on call records, is a significant challenge due to the complexity and variability of the data. Despite the efforts, no adequate studies were found in the literature on this subject. The development of a predictive model for personality profiling of mobile network customers based on information from the BI data warehouse works on models trained on Android system data, therefore seems to be a unique undertaking. The main challenge becomes adapting the warehouse data to Android's data format and features, which is a task that requires significant amount of pre-processing.

Previous research efforts in the field of profiling data have focused on social media, using data from written interactions and social media reactivity to create personality profiles, usually presented as a set of discrete binary variables. This approach may not accurately reflect the nuances of human behaviours, which is more diverse and complex than binary classifications suggest.

Jha et al. (2018) work on a cross-property deep transfer learning framework demonstrates effective predictive modeling on small datasets by leveraging larger datasets, an approach that could be adapted for transferring personality profiles from extensive mobile data to more focused BI systems.

Hanczar et al. (2022) assessed deep learning and transfer learning for cancer prediction using gene expression data, revealing the potential of transfer learning between different conditions, a concept applicable to the varied nature of mobile and BI data.

Sarmas et al. (2022) discussed transfer learning strategies for solar power forecasting under data scarcity, illustrating the significant improvement in predictive accuracy with limited data, a scenario akin to BI systems utilizing sparse call-log data. A theoretical and practical approach to transfer learning across datasets with different input dimensions, as proposed in a study focusing on linear regression, provides a robust framework for adapting mobile-derived personality profiles to the differing structure of BI system data.

In the context of existing literature, a meta-analysis of studies examining the relationship between smartphone data and the Big Five personality traits found that extraversion showed the greatest association with digital footprints obtained from smartphone and social media data. Which would perhaps indicate the potential of using data from call and text logs to better predict extraversion, as opposed to other types of smartphone data.

When it comes to knowledge transfer methodologies between data sets of different characteristics and sizes, the concept of transfer learning in AI and ML is increasingly used in various fields. For example, in materials science, cross-property deep transfer learning frameworks have been developed to use models trained on large data-sets to create models on smaller data-sets with different properties for improved material detection (Gupta et al., 2021). This approach enables the development of robust and accurate models on small data-sets, where labeled larger data-sets may not be readily available. Similarly, in network biology, transfer learning is used to leverage deep learning models pre-trained on large-scale general data-sets to adapt to tasks with limited task-specific data, such as disease modelling with limited patient data (Theodoris et al., 2023). This method has been shown to increase predictive accuracy and accelerate the discovery of key network regulators and therapeutic targets.

These examples from materials science and network biology demonstrate the feasibility and effectiveness of transfer learning methodologies in addressing the challenges posed by predictive modelling using data-sets of varying sizes and characteristics. This approach, although different from the research conducted on the integration of predictive models from mobile data with BI systems for the purposes of personality profiling, indicates the validity of searching for and researching such methods, where the key challenge is the variability of the data on which the model is trained and used.

4. Research Method

The overall methodology chosen to carry out the required research is Design Science Hevner and Chateerjee (2012). Following the Design Science framework, the presented research consists of the following steps presented in Fig. 1 Design Science Research can be defined as a six-step procedure proposed by Peffers et al. (2007) and specified as a sequence composed of 3 iterative cycles: relevance, rigor and design.

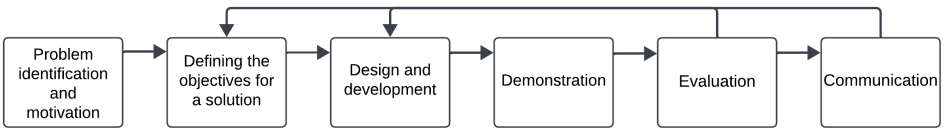

Figure 1: Design Science Research Methodology Process

According to the Design Science methodology, the research process aims to create artifacts that are verified for usefulness in the evaluation process in a given environment using the Framework of Evaluation for Design Science (FEDS) described by Venable et al. (2016). The research process in design science begins with problem identification, which involves defining research challenges and assessing their potential usefulness. This is followed by iterative verification against the existing knowledge base, known as the relevance cycle, and ongoing supervision, referred to as the rigor cycle. Within this iterative framework, the expected goals for the artifact are identified, and the method for its creation is designed. The process culminates in a demonstration of the created artifact, showcasing how it solves the problem within the appropriate context. Subsequently, the solution is evaluated in order to measure its effectiveness. The final stage involves communicating the results to relevant stakeholders.

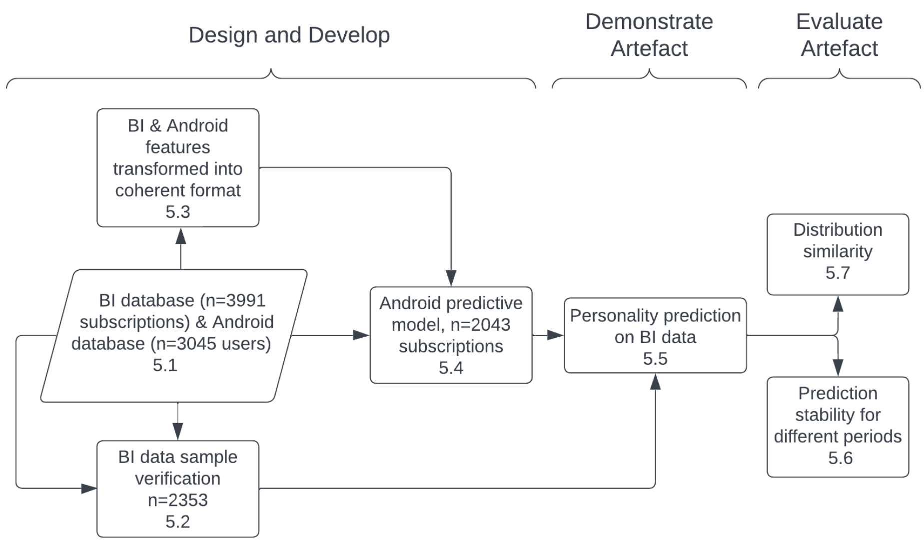

Figure 2: Design and development phase. Detailed description in article - in the 5th section with subsections

5. Developing the Data Driven Model (design and development stage of the Design Science Research)

After identifying objectives, the research process was designed using the methodology described in the previous section, with the design and development, demonstration, and evaluation of the artefact stages separated. Our instance of the generic process diagram is shown in Fig.2, with numerical references to the chapters that describe the work performed. Transferring a predictive model based on one data source to data from another seemed feasible if a standard set of explanatory variables could be found in both databases or the data could be transformed to create a standard set. An additional advantageous condition would be the possibility of identifying a comparable population in terms of demographic characteristics in both sources - e.g. customers from different generations may differ significantly in how they use mobile telephony. These stages of preparatory work are described in chapters 5.1 to 5.4.

The differences between primary and secondary data are crucial from the point of view of the research problem related to model transfer. The starting point for developing the predictive model in this study was the knowledge from the work of Krzeminska (2022) on predicting personality based on mobile phone data. The Android OS-based predictive model used several categories of data collected by a smartphone application, which created one-time statistics on phone use (one-shot view). These statistics included, among others, information about the list of installed applications, call logs, contact lists, SMS logs, photos and information about system settings. A reference to a detailed description of this study and the stages of model creation was provided. As a result, a set of models was created for each of the five characteristics and the models had different accuracy and F1 scores for each of them. The highest accuracy and F1 score were obtained for Extraversion (0.79 and 0.76) and the lowest for Agreeableness (0.70 and 0.58, respectively).

Ideally, it would be possible to annotate customer data from the BI warehouse with the results of personality surveys, but this is an impossible task due to costs, time and the unwillingness of respondents to complete surveys. To illustrate, paid recruitment of subjects when creating a model on Android data had a response rate below 10%. Although the highest quality metrics were achieved by models calculating a personality profile from all categories of data collected from a smartphone, a new model had to be created to transfer the Android model to the telco warehouse based only on a small part of the smartphone statistics, namely on the call logs. Such a significant data reduction means reducing predictive features by up to 95 percent.

5.1. Comparison of data-sets: Smartphone data vs BI telco warehouse data set

As already mentioned, the first step in implementing the model transfer was to extract a set of variables and features based on them from both data sources. The comparison concerned a set of smartphone statistics collected from 3,045 users of phone models with the Android operating system (input set) and a randomly selected sample of data from the BI warehouse from 3,991 unique mobile phone numbers (target set). In both cases, the data, although anonymous, came from people who consented to the processing of their data in order to create statistics. The training set (smartphones) included broad and diverse types of information. From the elementary categories, such as the battery level, to the more sophisticated, such as the frequency of emoji use in text messages. The users were customers of 4 major companies in the Polish telecommunications market (Orange, T-mobile, Plus, Play). Demographic characteristics were also assigned to the conversation logs: age, gender and the classification of 5 personality traits of the users (openness, conscientiousness, extraversion, agreeableness, neuroticism). The Big Five classification was based on the self-assessment results in a standardized questionnaire and included three classes for each feature. Classes were created based on a standardized sten scale: the low level was 1-3 sten, the medium level was 4-7 sten, and the high level was 8-10 sten. (after (Krzeminska, 2022)).

In the first step, the scope of input data was limited to statistics and features from call logs, which reduced the size of the database to 2,043 people with any calls registered during last 18 months. The data in the set included: *statistics from the call history and concerned tags, call duration, call types (incoming, outgoing, missed, rejected) and the types of telephone between which the call took place (mobile, landline)*.

The target set is a sample of randomly selected accounts active in 2023, downloaded from the Orange Polska telecommunications data warehouse (3,991 unique mobile phone numbers). Since the data from the warehouse is stripped of the exact connection timestamp, records from given day are aggregated in 24-hour periods. In order to best harmonize the data from the warehouse with the data from the input database, the following selection criteria were adopted in terms of user characteristics:
- payer accounts needed to have at least one active mobile phone
- data for this number should have at least one call dialled, received or unanswered

Finally, in the target database, call statistics for a given day included: *the total number of incoming and outgoing calls, also divided into on-net and off-net categories, corresponding to Orange's own network and that of other telecommunications companies; a total number of received, dialled and missed calls; daily duration of incoming and outgoing calls in minutes (also divided into on-net and off-net categories), daily duration of roaming and international calls (sum of incoming and outgoing calls) in minutes, duration of incoming calls in three periods - in the morning (00-12), in the afternoon (12-5 p.m.) and in the evening (5-12 p.m.); information about the number from which most calls are made to and from for a given number*.

5.2. BI data sample verification and adjusting

The data warehouse contains data for the entire Orange B2C customer population and differs from the data sample collected by the Dr Character application in two ways.

I. It is broader:
- contains data from users of both Android and IOS operating systems
- contains data from all subscriptions and therefore more than one phone for some subscribers
- contains subscriptions described as one-person businesses (which are not a typical individual user)

II. It is narrower:
- contains data of Orange customers, it does not include customers of the other three main companies in the Polish telco market (although Orange customers may have services from other companies - this information is not available in the BI database)
- does not include data on the use of applications on smartphones, which is available for customers from the Android database. Combining call logs with information on communication taking place via applications would give a more complete picture of users behaviour.

- does not provide information about timestamp of each call

To ensure maximum consistency between BI data sample and Android database, the following steps were taken: removing cases where only international calls are performed, excluding one-person businesses from sample and filtering out multi-subscriptions (phone numbers linked to single account). This procedure reduced the sample from 3991 to 2353 telephone numbers.

At the time of the study, there was no possibility of selecting only users with Android system smartphones for BI dataset. Another difference between the collections that has not been resolved is the issue of full telco market representation in Android dataset versus Orange users only in BI dataset. Restricting Android dataset to Orange customers only would reduce it to a quarter of its original size and would seriously impair possibility of machine learning training.

For both databases, acquisition of knowledge about gender and age of subscriptions users is possible. However, in case of BI sample data set, it is only available partially for two reasons. Firstly, there is no data for 38.1% of billing owners in terms of gender and no data for 39.4% of owners in terms of year of birth. Furthermore, there is also no absolute certainty about demographic characteristics of exact users of each subscription. Under Polish law, the billing owner may have multiple mobile phone numbers assigned to him or her, which may be used by the owner or other people (for example, family members). It may only be assumed that single subscriptions billing owner demographics is the correct description of the actual user. Due to the above caveats, gender and age were not used as variables in the predictive modelling.

5.3. Android and BI features transformed into coherent format

Both databases contain variables carrying information on the number of telephone calls made or received. However, they are aggregated at a different level and require transformations in the set with a higher level of detail. In addition, in the warehouse data, calls are separated into those made within and outside the Orange network. These need to be combined, as this type of information is not available at the Android level.

This diversity of datasets meant that both had to be modified to create a set of common features for predictive modelling. In the first dataset, the call histories from the timestamps were aggregated to the daily level, and converted to the form found in the data warehouse. The common variables on which the predictive model could then be built were the number of outgoing, incoming and missed calls, as well as the length of incoming calls in the three time intervals. The average time of outgoing and incoming calls per day for both bases was also calculated.

The Python *pandas* and *numpy* libraries, commonly used in Data Science, were used for the transformations. The final result of this step was to produce two time series databases with an identical set of variables, albeit from different sources: *total number of calls, number of incoming calls, number of outgoing calls, summed duration of incoming calls in hours 00-12, summed duration of incoming calls in hours 12-17, summed duration of incoming calls in hours 17-00, total duration of incoming calls, total duration of outgoing calls, mean duration of incoming calls, mean duration of outgoing calls, day of week, working/weekend day (binary), day with/without calls (binary).*

5.4. Android predictive model

The set of call logs prepared for the machine learning modelling consisted of daily statistics in a format corresponding to the BI data. Depending on the number of days per month with calls made or received ("active days"), the sample size of users with data changes. The higher the average number of "active days" per month, the fewer cases are available for machine learning training.

Figure 3 shows the relationship between the minimum available number of days with calls per month and the sample size. Due to the possible relationship between these two dimensions and the prediction results, several different cut-offs were made for experimental purposes and for comparison (5, 10, 15, 20 and 25 or more "active days" per month).

The first step in preparing predictive models was to use the Python library *tsfresh*, which operates on time-series data. It produces hundreds of potentially important predictor variables in an automated way and then tests their significance against each of the user characteristics targeted for prediction. The elimination of less important variables is then performed and the final set of predictive features is obtained - in the form of statistics describing the time series. Examples of the resulting variables range from the simplest, such as the sum of values, to the median, to more complex ones such as the permutation of entropy in the time series. With data set covering 6 months of call logs and 5 cut-offs of "active days", it was possible to create 6x5 matrix with 30 alternative data sets. Each one had different level of activity of users and different length of history of call logs. It turned out that the *tsfresh* library was able to

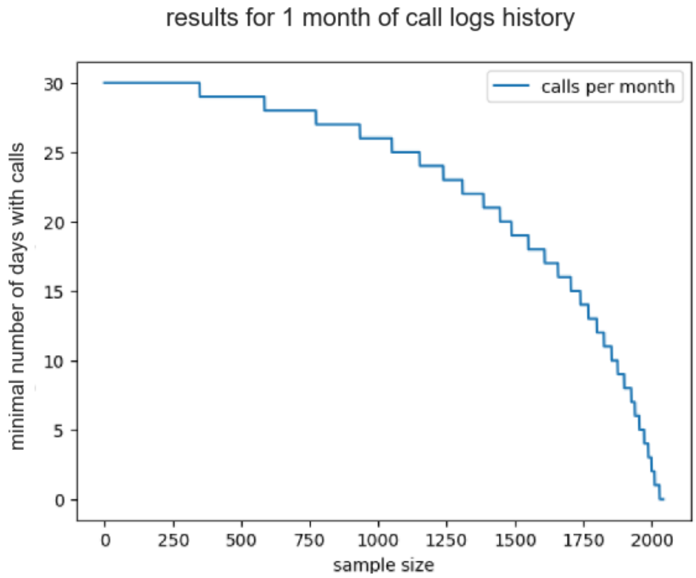

Figure 3: Minimum available number of days with calls per month versus sample size

create significant features for all data sets in matrix only with Extraversion trait. For example, for users with 1 month of calls history and at least 5 calls per month, features created with that library were not significant in predicting Conscientiousness. But with adding more data points (another month of calls history) or limiting data set to users with more informative data (at least 10 calls per month) - features created were significant in predicting Conscientiousness. Finally, Conscientiousness received 24 and Agreeableness 16 lists of features for corresponding data sets. There were no significant features created by tsfresh for Neuroticism and Openness traits.

The second step in the process was to investigate performance of 30 popular classifiers from *scikit-learn*, *xgboost* and *lightgbm* Python libraries. The algorithms were trained on a training set and tested on a test set. The measures used to check the quality of the algorithms were accuracy and F1 score. A full list of the applied algorithms is provided in the Appendix.

The final results obtained during experiments are presented at Table 1. The best performing algorithms were (depending on the number of months and number of "active days"): XGBClassifier, LogisticRegression, AdaBoost-Classifier, ExtraTreesClassifier, BaggingClassifier and RandomForestClassifier.

5.5. Personality Prediction on BI

The previous stage completed the data processing phase for customers' smartphones. Predictive models were developed for various combinations of call history lengths and different customer call intensity groups. The models that performed best were then used on data from the data warehouse, which, as a reminder, does not include direct information about users' personality profiles. This constraint necessitates innovative methods to verify predictions on unlabeled data. Three distinct methods were identified to validate the quality of the predictive models transferred from the smartphone environment to the data warehouse, each varying in execution complexity.

Method 1: Distribution Comparison. The initial approach involves comparing the distributions of the predicted personality traits between the results for the smartphone-derived model and the BI warehouse model. It is particularly important to assess whether both models produce predictions from the same probability distribution for personality trait classes. This comparison helps determine if the transferred model maintains similar predictive behavior across different data sources.

Method 2: Temporal Consistency. The second method examines the consistency of predictions for the same customers over different time periods, namely across six consecutive months. Since personality traits are generally stable over time, consistent predictions would suggest the model's reliability. This approach provides a robust measure

TECHNIUM
Technium Vol. 23, pp.6-17 (2024)
ISSN: 2668-778X
www.techniumscience.com

EXTRAVERSION		Accuracy					F1				
		average number of days with calls per month									
		5	10	15	20	25	5	10	15	20	25
number of months with calls available	1	0.67	**0.70**	**0.70**	**0.69**	0.68	**0.58**	**0.61**	**0.60**	**0.60**	**0.57**
	2	**0.68**	0.67	**0.69**	**0.69**	**0.70**	**0.56**	**0.59**	**0.57**	**0.61**	**0.64**
	3	**0.69**	**0.69**	0.68	**0.69**	**0.71**	**0.63**	**0.59**	**0.59**	**0.57**	**0.62**
	4	**0.69**	**0.71**	0.67	**0.70**	**0.70**	**0.59**	**0.66**	**0.60**	**0.59**	**0.65**
	5	**0.68**	0.67	0.68	**0.69**	**0.72**	**0.60**	**0.60**	**0.58**	**0.60**	**0.65**
	6	**0.69**	0.67	0.68	**0.71**	0.68	**0.59**	**0.61**	**0.62**	**0.62**	**0.61**

CONSCIENTIOUSNESS		Accuracy					F1				
		average number of days with calls per month									
		5	10	15	20	25	5	10	15	20	25
number of months with calls available	1	-	**0.70**	0.69	0.69	-	-	**0.63**	0.56	**0.58**	-
	2	**0.71**	**0.70**	**0.70**	0.69	0.68	**0.62**	**0.60**	**0.59**	0.56	0.55
	3	**0.70**	**0.71**	**0.70**	**0.71**	-	**0.59**	**0.61**	**0.59**	**0.62**	-
	4	**0.71**	**0.71**	**0.70**	**0.71**	-	**0.63**	**0.61**	**0.62**	**0.60**	-
	5	**0.70**	**0.71**	**0.72**	**0.70**	-	**0.62**	**0.61**	**0.63**	**0.60**	-
	6	**0.70**	**0.73**	**0.71**	**0.70**	-	**0.60**	**0.64**	**0.61**	**0.59**	-

AGREEABLENESS		Accuracy					F1				
		average number of days with calls per month									
		5	10	15	20	25	5	10	15	20	25
number of months with calls available	1	-	-	-	-	-	-	-	-	-	-
	2	0.70	**0.72**	**0.72**	-	-	**0.58**	**0.61**	**0.6**	-	-
	3	**0.72**	**0.73**	**0.71**	-	-	**0.63**	**0.63**	**0.62**	-	-
	4	**0.71**	**0.72**	**0.71**	-	-	**0.61**	**0.61**	**0.64**	-	-
	5	**0.71**	**0.72**	**0.71**	-	-	**0.60**	**0.66**	**0.60**	-	-
	6	**0.71**	**0.72**	**0.71**	**0.69**	-	**0.63**	**0.65**	**0.61**	**0.59**	-

Table 1
Accuracy and F1 scores for Extraversion, Conscientiousness and Agreeableness trait for different amount of data available for individual.

	stability of predictions			similarity of distribution		
	minimum	maximum	mean	minimum	maximum	mean
Extraversion	68,8%	92,5%	84,0%	45.2%	99.2%	72.6%
Conscientiousness	76,0%	81,6%	79,4%	89.8%	98.9%	93.8%
Agreeableness	73.8%	97.4%	87.7%	56.4%	98.1%	84.1%

Table 2
Minimum, maximum, mean stability of prediction and similarity of distribution. The higher the result, the higher stability or similarity

of the temporal stability of the model's predictions, ensuring they reflect enduring personality traits rather than transient behaviors.

Method 3: Behavioral Correspondence. The third method involves correlating predicted personality traits with observed customer behavior. This method requires in-depth collaboration with business stakeholders to obtain detailed knowledge about marketing campaigns, internal processes, and actual customer behaviors. By examining the correspondence between predicted traits and real-life behavioral data, this method validates the practical applicability and accuracy of the personality predictions in reflecting customers' actual personalities and behaviors. The use of this method is planned in future studies.

The results of verification with Method 1 and Method 2 are presented in Table 2.

Technium Vol. 23, pp.6-17 (2024)
ISSN: 2668-778X
www.techniumscience.com

5.6. Predictions stability for different periods

The process of verifying the stability of predictions involved assessing how consistently the models assigned personality trait classes to the same set of customers over multiple time periods. To achieve this, the best-performing models trained on smartphone data were used to make predictions on BI warehouse data for five distinct months. The same set of customers was used for each month to ensure comparability.

The stability of predictions was quantified by measuring the percentage of customers who were assigned the same personality trait class across the different months. This stability metric could range from 0% (if no customer ever had the same class in the compared months) to 100% (if all customers consistently had the same class in the compared months).

For Extraversion, the stability of predictions varied from 68.8% to 92.5%, with the highest stability observed in models trained on data from groups with at least 10 calls per month. This suggests that Extraversion is relatively stable over time, particularly when sufficient data points are available. Conscientiousness exhibited stability ranging from 76.0% to 81.6%, with the highest stability found in models trained on data from groups with at least 20 calls per month. This indicates a moderate level of temporal consistency for Conscientiousness predictions. Lastly, Agreeableness showed stability ranging from 73.8% to 97.4%, with varying results based on the intensity of call activity. This variability highlights the potential influence of data quantity and quality on the stability of predictions.

Overall, the stability of predictions across different periods underscores the reliability of the transferred models in maintaining consistent personality trait classifications over time. This temporal consistency reinforces the validity of using call log data for personality prediction within BI systems, provided that sufficient and representative data are available.

5.7. Distribution similarity

To verify the similarity of distributions between the smartphone-derived model and the BI data model, the following method was adopted. First, a personality prediction was made on a random test set from the smartphone data using the models trained previously. Then, the same model was used to predict personality traits on a random test set from the BI warehouse data. The next step involved comparing the results using the Kolmogorov-Smirnov (K-S) test, a statistical test used to determine if two unidimensional probability distributions differ. The more similar the distributions of both samples, the more reliable the predictions delivered by the models.

The application of the K-S test indicated varying degrees of similarity between the distributions of predicted personality traits from the smartphone data and the BI data. For Extraversion, the similarity ranged from 45.2% to 99.2%, with a mean similarity of 72.6%. This suggests that while there are instances of high congruence, the variability is notable. For Conscientiousness, the similarity ranged from 89.8% to 98.9%, with a mean similarity of 93.8%, indicating a high level of consistency between the two data sources. Agreeableness showed a similarity range from 56.4% to 98.1%, with a mean of 84.1%, reflecting moderate to high congruence.

These results demonstrate that the predictive models can produce similar probability distributions across different data sources, especially for traits like Conscientiousness, where the similarity is particularly high. This high degree of similarity reinforces the feasibility of transferring predictive models from smartphone data to BI systems, though the variability seen in traits like Extraversion and Agreeableness suggests that further refinement and calibration of the models may be necessary for these traits.

In conclusion, the use of the Kolmogorov-Smirnov test to compare probability distributions has provided a robust means of validating the transferred predictive models. The findings support the potential of integrating personality insights derived from smartphone data into BI systems, contributing to the development of more personalized and effective business intelligence applications. Future research should continue to explore methods to enhance the alignment of predictive models across diverse data sources, ensuring consistent and reliable personality predictions.

6. Discussion

The integration of predictive personality models derived from smartphone data into Business Intelligence (BI) systems is an innovative approach that bridges the gap between individual-level data insights and large-scale business applications. This study successfully demonstrates that personality traits such as Extraversion, Conscientiousness and Agreeableness can be predicted from call logs data within BI systems with features generated through *tsfresh* library. The distribution comparison and temporal consistency validation methods used provide a robust framework for assessing the accuracy and reliability of these predictions.

Recent studies in personality prediction have explored various data sources and methodologies, highlighting both the potential and the challenges of such endeavors. For example, research utilizing natural language processing (NLP) has effectively predicted personality traits from social media texts by leveraging large datasets and sophisticated machine learning models (Jang et al., 2022; Mehta et al., 2020). These studies emphasize the value of diverse data sources and complex algorithms in improving prediction accuracy.

In article Jang et al. (2022), deep learning and transfer learning techniques have shown to enhance predictive performance, particularly in contexts with limited data. This aligns with our study's approach of transferring models trained on smartphone data to a BI environment, illustrating the applicability of cross-domain learning strategies. Additionally, the use of psycholinguistic features extracted from text data has provided insights into the cognitive and emotional dimensions of personality (Mehta et al., 2020). Although our study focuses on call log data, these findings suggest that incorporating multiple data types, such as text and call logs, could further enhance the robustness and comprehensiveness of personality predictions.

Our previous research discussed in Krzeminska and Rzeznik (2021) indicates how digital data can be utilized to profile users and adapt services accordingly, which parallels our focus on leveraging call log data for personality prediction. Both studies emphasize the importance of personalizing user experiences based on personality insights derived from digital footprints.

Future research should strive for integrating multimodal data, combining call logs with other data types like text messages, social media interactions, and app usage, to provide a more holistic view of user behavior and improve prediction accuracy. The exploration of advanced machine learning techniques, including deep learning models and ensemble methods, could further enhance the predictive power and generalizability of the models across different datasets. Different methods of creating predictive variables such as manual feature engineering, should make it possible to predict the other personality traits in the Big 5 model (Openness, Neuroticism) from call logs data.

Developing models capable of real-time personality predictions and dynamic adaptation to new data is crucial for practical applications in business environments. Addressing ethical and privacy considerations is also essential; future studies should focus on creating frameworks and guidelines that protect user data while maximizing the benefits of personality insights. Conducting longitudinal studies to track changes in personality predictions over time will help validate the stability and reliability of the models, providing deeper insights into the temporal dynamics of personality traits.

7. Conclusion

This study aimed to determine the feasibility of transferring a personality prediction model, initially trained on smartphone data, to a Business Intelligence (BI) system utilizing call logs. Our study addressed two primary questions: whether profiles can be predicted using models trained on different data sources and how to measure the quality of such forecasts.

This study demonstrates that profiles can be predicted using models trained on different data sources. Specifically, a personality prediction model trained on diverse smartphone data was successfully transferred to a Business Intelligence (BI) system utilizing call logs. The findings indicate that personality traits such as Extraversion, Conscientiousnes and Agreeableness can be reliably predicted from BI call log data through meticulous data preprocessing and feature engineering (RQ1).

To measure the quality of these forecasts (RQ2), distribution comparison and temporal consistency metrics were employed. Distribution comparison involved using the Kolmogorov-Smirnov test to compare the distributions of predicted traits from both the smartphone-derived model and the BI model, ensuring their similarity. Temporal consistency assessed the stability of predictions for the same customers over six months, confirming that personality traits remained stable over time. Although a third method, behavioral correspondence, was proposed, it was not utilized in this research. These validation methods confirmed the robustness and reliability of the transferred models within the BI context, demonstrating their practical applicability and potential for enhancing business intelligence systems.

The clear business benefits can be identified in this research experiments. By enabling the prediction of personality traits from call logs, businesses can enhance their personalization strategies, tailoring services and interactions to individual customers more effectively. This approach is scalable, as demonstrated by the successful transfer of the model from a smaller smartphone dataset to a larger BI dataset. Moreover, the research emphasized the importance of ethical considerations, providing a responsible framework for using personality data in business applications without gathering additional personal data from users. The experiment is an interesting case of re-using an existing model predicting a feature in order to generate similar information based only on statistical analyses without the need to collect additional data.

Addressing ethical and privacy concerns remains essential, and future studies should focus on creating frameworks that protect user data while maximizing the benefits of personality insights. Conducting longitudinal studies to track changes in personality predictions over time will help validate the models' stability and reliability, providing deeper insights into the temporal dynamics of personality traits. By addressing the above-mentioned future research areas, such as integrating multi modal data with call logs or developing real-time prediction models, future studies can build upon the foundations laid by this work, advancing the field of personality prediction and its applications in business intelligence and beyond. This research contributes to the growing body of knowledge on leveraging digital traces for personality assessment, paving the way for more personalized and effective business solutions.

This research significantly contributes to the domain of data science and business intelligence by demonstrating the feasibility of transferring personality prediction models across disparate data sources. By successfully adapting a model trained on diverse smartphone data to predict personality traits from BI call log data, this study bridges the gap between individual-level digital behavior analysis and large-scale business applications. The research highlights the effectiveness of meticulous data preprocessing, feature engineering, and innovative validation methods, such as distribution comparison and temporal consistency, in ensuring the robustness and reliability of transferred models. Furthermore, this study underscores the practical implications of integrating personality insights into BI systems, paving the way for more personalized and effective business strategies while addressing ethical considerations and data privacy challenges. This work not only advances the technical methodology of model transfer but also expands the application of personality prediction in commercial contexts, offering a scalable approach to enhancing customer understanding and personalization.

Acknowledgements

The experiment described in this paper was financed by the Orange Innovation Research and B2C Customer Journey departments in Orange and was technically prepared by the AI Skills Center department in Orange Innovation Poland under Paweł Tuszyński's lead. The authors wish to thank Michał Szczerbak for supervision.

References

Berkovsky, S., Taib, R., Koprinska, I., Wang, E., Zeng, Y., Li, J., Kleitman, S., 2019. Detecting personality traits using eye-tracking data, in: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, pp. 1–12.

Bin Tareaf, R., Alhosseini, S.A., Meinel, C., 2019. Facial based personality prediction models for estimating individuals private traits, in: Proceedings - 2019 IEEE Intl Conf on Parallel and Distributed Processing with Applications, Big Data and Cloud Computing, Sustainable Computing and Communications, Social Computing and Networking, ISPA,BDCloud,SustainCom,SocialCom 2019. doi:10.1109ISPA-BDCloud-SustainCom-SocialCom48970.2019.00233.

Brdulak, H., 2014. Poszukiwanie nowych paradygmatów w zarządzaniu łańcuchem dostaw w warunkach stagnacji gospodarczej. Zeszyty Naukowe Uniwersytetu Gdańskiego. Ekonomika Transportu i Logistyka , 23–36.

Costa Jr, P.T., McCrae, R.R., 1992. Four ways five factors are basic. Personality and individual differences 13, 653–665.

DeYoung, C.G., 2015. Cybernetic big five theory. Journal of research in personality 56, 33–58.

Gupta, V., Choudhary, K., Tavazza, F., Campbell, C., Liao, W.k., Choudhary, A., Agrawal, A., 2021. Cross-property deep transfer learning framework for enhanced predictive analytics on small materials data. Nature communications 12, 6595.

Hanczar, B., Bourgeais, V., Zehraoui, F., 2022. Assessment of deep learning and transfer learning for cancer prediction based on gene expression data. BMC bioinformatics 23, 262.

Hevner, A., Chateerjee, S., 2012. Integrated Series in Information Systems Volume 28 doi:10.1007/978-1-4419-6108-2.

Jang, J., Yoon, S., Son, G., Kang, M., Choeh, J.Y., Choi, K.H., 2022. Predicting personality and psychological distress using natural language processing: a study protocol. Frontiers in Psychology 13, 865541.

Jha, D., Ward, L., Paul, A., Liao, W.k., Choudhary, A., Wolverton, C., Agrawal, A., 2018. Elemnet: Deep learning the chemistry of materials from only elemental composition. Scientific reports 8, 17593.

Kalimeri, K., Lepri, B., Pianesi, F., 2013. Going beyond traits: Multimodal classification of personality states in the wild, in: ICMI 2013 - Proceedings of the 2013 ACM International Conference on Multimodal Interaction, pp. 27–34. doi:10.1145/2522848.2522878.

Khan, A.S., Ahmad, H., Asghar, M.Z., Saddozai, F.K., Arif, A., Khalid, H.A., 2020. Personality classification from online text using machine learning approach. International Journal of Advanced Computer Science and Applications 11. doi:10.14569/ijacsa.2020.0110358.

Kosinski, M., Matz, S.C., Gosling, S.D., Popov, V., Stillwell, D., 2015. Facebook as a research tool for the social sciences: Opportunities, challenges, ethical considerations, and practical guidelines. American psychologist 70, 543.

Krzeminska, I., 2022. Automatic data-based personality assessment as a method of electronic services auto-personalisation (phd thesis). https://bip.ue.poznan.pl/download/attachment/1325/rozprawa-doktorska-mgr-i-krzeminskiej.pdf.

Krzeminska, I., Rzeznik, J., 2021. Personality-based lexical differences in services adaptation process. Technium: Romanian Journal of Applied Sciences and Technology 3, 61–73. URL: https://techniumscience.com/index.php/technium/article/view/2101.

Krzeminska, I., Szmydt, M., 2021. Personality based data-driven personalization as an integral part of the mobile application, in: International Conference on Business Information Systems, Springer. pp. 144–155.

Liu, L., Preoţiuc-Pietro, D., Samani, Z.R., Moghaddam, M.E., Ungar, L., 2016. Analyzing personality through social media profile picture choice. Proceedings of the 10th International Conference on Web and Social Media, ICWSM 2016 , 211–220.

Matz, S.C., Kosinski, M., Nave, G., Stillwell, D.J., 2017. Psychological targeting as an effective approach to digital mass persuasion. Proceedings of the National Academy of Sciences of the USA 114, 12714–12719. doi:10.1073/pnas.1710966114.

Mehta, Y., Fatehi, S., Kazameini, A., Stachl, C., Cambria, E., Eetemadi, S., 2020. Bottom-up and top-down: Predicting personality with psycholinguistic and language model features , 1184–1189doi:10.1109/ICDM50108.2020.00146.

Montjoye, Y.a.D., Quoidbach, J., Robic, F., 2013. Predicting personality using novel phone-based metrics. Social Computing, Behavioral-Cultural Modeling and Prediction Lecture Notes in Computer Science , 48–55.

Ning, H., Dhelim, S., Aung, N., 2019. Personet: Friend recommendation system based on big-five personality traits and hybrid filtering. IEEE Transactions on Computational Social Systems 6. doi:10.1109/TCSS.2019.2903857.

Peffers, K., Tuunanen, T., Rothenberger, M.A., Chatterjee, S., 2007. A design science research methodology for information systems research. Journal of management information systems 24, 45–77.

Sarmas, E., Dimitropoulos, N., Marinakis, V., Mylona, Z., Doukas, H., 2022. Transfer learning strategies for solar power forecasting under data scarcity. Scientific Reports 12, 14643.

Stajner, S., Yenikent, S., 2021. Why is MBTI personality detection from texts a difficult task?, in: Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume, Association for Computational Linguistics, Online. pp. 3580–3589. URL: https://www.aclweb.org/anthology/2021.eacl-main.312.

Swaminathan, V., 2003. The impact of recommendation agents on consumer evaluation and choice: the moderating role of category risk, product complexity, and consumer knowledge. Journal of Consumer Psychology 13, 93–101.

Theodoris, C.V., Xiao, L., Chopra, A., Chaffin, M.D., Al Sayed, Z.R., Hill, M.C., Mantineo, H., Brydon, E.M., Zeng, Z., Liu, X.S., et al., 2023. Transfer learning enables predictions in network biology. Nature , 1–9.

Venable, J., Pries-Heje, J., Baskerville, R., 2016. Feds: a framework for evaluation in design science research. European journal of information systems 25, 77–89.

Xu, R., Frey, R.M., Ilic, A., 2016. Individual Differences and Mobile Service Adoption: An Empirical Analysis, in: Proceedings - 2016 IEEE 2nd International Conference on Big Data Computing Service and Applications, BigDataService 2016, Institute of Electrical and Electronics Engineers Inc.. pp. 234–243. doi:10.1109/BigDataService.2016.15.

Appendix

List of predictive algorithms used (in alphabetical order):

AdaBoostClassifier, BaggingClassifier, BernoulliNB, CalibratedClassifierCV, DecisionTreeClassifier, ExtraTreesClassifier, GaussianNB, GradientBoostingClassifier, HistGradientBoostingClassifier, KNeighborsClassifier, LabelPropagation, LabelSpreading, LinearDiscriminantAnalysis, LinearSVC, LogisticRegression, LogisticRegressionCV, MLPClassifier, NearestCentroid, NuSVC, PassiveAggressiveClassifier, Perceptron, QuadraticDiscriminantAnalysis, RandomForestClassifier, RidgeClassifier, RidgeClassifierCV, SGDClassifier, SVC

Jakub Rzeznik is an expert in Orange Innovation Poland and PhD candidate in Warsaw School of Economics. He received his M. Sc in sociology from Institute of Sociology at the University of Warsaw. He has more than 15 years of experience in the field of data analysis, statistics and research - especially in solving classification and regression problems. He specializes in statistical modelling and machine learning techniques (supervised and unsupervised). He currently consults on R&D projects in the areas of mobile application data processing, segmentation using telecom operator Big Data and data-driven product development.

Izabella-Krzemińska is an expert in Orange Innovation Poland. She received her M. Sc in psychology and psychometric from University of Warsaw and PhD in Data Science in Business in the Faculty of Computer Science and Electronic Economy/Poznan University of Economics. She is a researcher and data analyst with almost 20 years of work experience (both research institutes, consulting and client-side). She manages R&D projects concerning using operators data for commercial purposes, creating new services based on data (geolocation) and creating additional value based on digital footprint, especially in the psychology area (cyber-personality).