24th European Conference on Artificial Intelligence - ECAI 2020
Santiago de Compostela, Spain

MobiSenseUs: Inferring Aggregate Objective and Subjective Well-being from Mobile Data

Martin Hillebrand and **Imran Khan** and **Filipa Peleja** and **Nuria Oliver** [1]

Abstract. Assessing the well-being of a population is of utmost importance for policy and decision makers so they can design appropriate policies and interventions to improve the quality of life of their citizens. Traditional methods to determine aggregate well-being consist of surveys which are expensive to obtain and difficult to scale. Thanks to the availability of large-scale human behavioral data, new methods to assess well-being might be possible. In this paper we describe one of such methods: MobiSenseUs, a machine-learning based system to automatically estimate geographically aggregated objective and subjective well-being measures in the UK from mobile data. We propose a comprehensive battery of features that capture different aspects of human behavior – *i.e.* communication patterns, mobile app usage and spatial mobility – from two sources of pseudonymized mobile data of more than one million smartphone users. We are the first to build machine-learning models to predict both objective (IMD) and subjective (SWB) indicators in the UK from these mobile features. We find that the IMD can be predicted more accurately than SWB, reaching **99%** and **78%** average accuracies in a binary classification task for the IMD and SWB, respectively. We analyze the most predictive features and derive implications for the design of data-driven machine-learning public health policy systems.

1 Introduction and Motivation

Well-being is a complex, multi-dimensional construct which is typically split into two domains: objective and subjective well-being[2].

Objective well-being is assessed using indicators that measure the state of a region from an economy, education, safety, physical/infrastructure, employment, etc. perspective. Objective well-being tends to capture the general state of a large group of individuals who live in the same region rather than individual perspectives. It is based on tangible, quantitative and material indicators. In the UK, objective well-being is measured by the Index of Multiple Deprivation (IMD).

Subjective well-being (SWB) is characterized by an individual's internal subjective assessment, based on cognitive judgments and affective reactions to their life experience. The concept was introduced by Diener [8] who proposed that a happy person emerges from his/her age, education, health, employment, social connections and other aspects, which *"call people to recognize and to live in accordance with their daimon or true self"* [35].

From a public policy perspective, assessing the well-being of a population is of utmost importance for policy and decision makers so they can design appropriate policies and interventions aimed at improving the quality of life of their citizens. Traditional methods to determine aggregate well-being consist of surveys carried out by the National Statistics Offices (NSOs), which are expensive to obtain, difficult to scale and subject to human error. Hence, the NSOs of most developed countries only carry out such surveys every few years.

Today we have access to large-scale human behavioral data that might be helpful to automatically assess a population's well-being. In this paper, we tackle such a challenge with MobiSensUs, a machine-learning based approach to infer objective and subjective aggregate well-being from two types of pseudonymized mobile data.

The paper is organized as follows. We first provide an overview of the most relevant published literature and our research questions in Section 2. We describe our data sources in Sections 3 and 4, followed by our results in Section 5. We conclude with a discussion of our findings and their implications in Section 6 and our conclusions in Section 7.

2 Related Work and Contributions

In recent years, researchers have been eager to understand the relationship between human behavioral data traces from *e.g.* social networks or mobile phone data and national statistics, including population counts, mobility and socio-economic levels. From the wide array of literature, we focus our literature review on published research that is aimed at inferring *spatially aggregated* –as opposed to individual– *well-being* indicators from passively collected human behavioral data sources, with an emphasis on mobile phone data. We structure the related work according to the type of behavioral data analyzed (non-mobile vs. mobile) and the type of well-being modeled (objective vs. subjective).

2.1 Well-being models from non-mobile phone data

Objective well-being: Several studies have aimed at predicting objective well-being in the UK, captured by the Index of Multiple Deprivation (IMD), from non-mobile phone data. The most commonly used data sources are Twitter [26], public transport card activity [17, 16, 30] and Foursquare [34, 27, 15]. The reported performances to estimate aggregate objective well-being (IMD) from these data sources are in the range of F1 = 0.7 for a binary classification between above-median vs below-median deprivation neighborhoods using state-of-the-art machine learning techniques. Existing works have limitations regarding the length of the study (just a few months) and the representativeness of the data, which we address in this paper.

Subjective well-being: Beyond objective well-being, we only found one study that tackled the prediction of spatially aggregated subjective well-being (SWB) from non-mobile data [21]. The authors of this study used the IMD and the Oyster Travel Card dataset to infer what is *needed*, combined with Navteq POI data to infer what is *offered* in more than 600 London neighbourhoods. Using all three data sources, they obtained an $R^2 = 0.25$ in a regression model with SWB as target variable.

[1] Work done while at Vodafone Research; email: nuria@alum.mit.edu

[2] See https://www.hsph.harvard.edu/health-happiness/research-new/positive-health/measurement-of-well-being/.

2.2 Well-being models from mobile phone data

In the literature, researchers have used two main types of large-scale passively collected human behavioral data from mobile phones: (1) Call Detail Records[3] (CDRs) which are collected by mobile network operators and register who called whom, when, where and for how long; and (2) data collected by smartphone applications which are running in the background.

Most previous work –if not all– has only used one type of mobile phone data alone or in combination with other non-mobile data sources (*e.g.* satellite data). The MobiSensUs project leverages both mobile data sources at the same time.

Objective well-being: Passively collected mobile phone usage data –and prominently Call Detail Records– has been analyzed in several studies to infer aggregate objective well-being and particularly aggregate socio-economic indicators.

Eagle *et al.* [10] showed that the IMD for England from 2004 is correlated with different aspects of call behavior (landline and mobile) as per their analysis of a CDR dataset collected in 2005. The more socio-economically deprived a region is, the larger the communication volume (time spent calling), the lower the number of contacts and the lower the entropy of calls (*i.e.* the higher the concentration of calls to/from a small number of contacts) of the individuals who live in that region.

The inference of objective well-being measures from mobile data in geographies other than the UK (and particularly developing economies) has also been carried out in several studies.

Pappalardo *et al.* [23] analyzed mobility patterns derived from CDRs in France and showed that municipalities where people have a higher mobility diversity (entropy of stop locations) are generally less socio-economically deprived. [32] and [13] showed that CDR-derived features (social network and mobility features) can be used to classify regions in a Latin American country with regards to their socio-economic level (three terciles, 80.7% accuracy). In a regression model using the same data, they obtain and adjusted $R^2 = 0.83$ with ordinary least squares regression. They also proposed in [12] a method to forecast future socio-economic levels at a state level in a Latino American country.

Several contributions submitted to the D4D Orange challenges in 2013 and 2015 showed that CDR-derived features can also be used to automatically infer regional differences in poverty in two African countries (Ivory Coast and Senegal), such as [31], [20] and [25] who built a poverty prediction model in Senegal with $R^2 = 0.672$.

Blumenstock *et al.* [3] joined mobile phone data and survey responses from 856 citizens in Rwanda to train a model of household wealth. They built estimates of household wealth at a district level with a very strong correlation ($r = 0.92$) with wealth indices from government survey data. Recently, Castillo *et al.* [4] analyzed CDRs in Ecuador (another developing country) and found that low communication volume is again indicative of poverty.

CDR data has also been used in conjunction with other data sources to predict measures of objective well-being.

Njugana and McSharry [22] combine features from CDRs with satellite data and information about population density into a regression model and predict the multi-dimensional poverty index in Rwanda with $R^2 = 0.76$, with population density being by far the most predictive regressor.

Steele *et al.* [33] also combine satellite data (vegetation, night-time lights, distance to urban areas) with CDRs and predict different poverty measures in Bangladesh. The best prediction was for the poverty facet *assets*, with $R^2 = 0.76$. Features computed using satellite data alone yield $R^2 = 0.74$. Conversely, the models aimed at predicting consumption and income only achieved $R^2 = 0.27$ each.

Features computed from CDRs and airtime credit purchases strongly correlate with measures collected via a food consumption survey and poverty indices and thus could serve to monitor food security [7]. CDRs combined with mobile money payment data can be used to accurately classify regions into one of five socio-economic status classes [11].

Interestingly, the findings in developing economies do not necessarily hold in developed countries. While communication volume positively correlates with socio-economic well-being in the context of a developing country like Ivory Coast [31], a small negative correlation has been found by [10] in a developed country like the UK. However similar to Ivory Coast, call entropy in the UK was found to be positively correlated with socio-economic well-being.

Finally, the exponential growth in adoption of smartphones has enabled additional research which has analyzed data captured by smartphones. However, to the best of our knowledge, we only found one large-scale aggregate study using smartphone sensors and well-being data [1], where the authors analyzed accelerometer data from users in 111 countries. This work found that countries with high inequalities in their level of physical activity are over proportionally countries with high obesity rates.

Subjective well-being: Despite a thorough literature research, we could not find any study using large-scale passively collected mobile data data to predict aggregate subjective well-being.

2.3 Contributions

Based on previous work, in this paper we address an existing gap in the literature which leads to five main contributions:

1. We build machine-learning models to automatically infer *both aggregate objective and subjective well-being* in the UK. We perform both classification and regression tasks and report performances that are at par or superior to the state-of-the-art;

2. We analyze *two sources of human behavioral mobile phone data*: CDRs and data collected by an Android app for a very large set of over 1 million users;

3. We design a battery of 16 human behavioral features that capture three important dimensions of human behavior: social and communication features, app usage features and mobility features. Several of our proposed features have never been used before for this purpose;

4. We carry out a feature importance analysis to understand how well suited are the different types of features to infer aggregate objective and subjective well-being; and

5. We derive three implications for the design of data-driven systems for public health policy making.

3 Measures of Well-Being in the UK

The UK's Index of Multiple Deprivation or IMD is the official measure of *objective well-being* in the UK. It is designed to determine the relative deprivation of small areas in the UK. It consists of a ranking of 32,844 small areas[4] from the most to the least deprived. It is composed of seven facets: Income, Employment, Health, Education, Barriers to Housing and Services, Crime and Living Environment. For a more detailed description of all facets we refer the readers to the IMD's technical report [29]. The IMD was published for the first time in 1998 and it has been updated five times since then in 2000, 2004, 2007, 2010 and in 2015. Given its high spatial granularity and differentiated profile, policy-makers and voluntary services have used

[3] https://en.wikipedia.org/wiki/Call$_{detail_{r}ecord}$

[4] Officially called Lower Layer Super Output Areas (LSOAs), which are geographic entities for the reporting of small area statistics in England and Wales with mean population of 1500.

the IMD to distribute funding or target resources with the intention to improve specific aspects of deprivation in the most deprived areas.

In 2010, the British Prime Minister at the time, David Cameron, launched the National Well-being Programme with the intention to extend traditional measures of objective well-being by applying a wider definition of well-being that also encompasses people's subjective feelings and experiences. Given that objective measures, such as the IMD, do not fully correspond with people's subjective experience, the ONS developed a *subjective well-being* (SWB) questionnaire where people are asked to directly evaluate their own lives. Since then, subjective well-being has been included in the evaluation procedure of a wide variety of different political areas, such as community learning, nature improvement, transportation or the health sector.

The SWB measure consists of four 10-point (0 - not at all; 10 - completely) survey questions about their Life Satisfaction, Worthwhile, Happiness and Anxiety. Life Satisfaction and Worthwhile represent the evaluative and eudemonic focus of subjective well-being, respectively, whereas Happiness and Anxiety capture facets of affective well-being. SWB answers are aggregated spatially at a local authority level by taking the mean of the respondents' individual values. The average values of SWB for the 2017 survey are 7.7 (Life Satisfaction), 7.9 (Worthwhile), 7.5 (Happiness) and 2.9 (Anxiety).

Given that we carry out our analysis on aggregate data, we are particularly interested in stable constructs with regards to subjective well-being, such as Life Satisfaction. Hence, our feature importance analysis mainly focuses on Life Satisfaction.

4 Mobile Phone Data Analysis

We model two types of target variables, the IMD and SWB, using two sources of mobile phone data. Here, we first describe our mobile phone data sources, followed by the spatial and temporal overlap between our target variables and the mobile phone data. Finally, we summarize our feature extraction process which yields 16 features in three categories: Social/Communication, App Usage and Mobility.

4.1 Mobile phone data

We use two types of mobile phone data, Call Detail Records (CDRs) and data collected by an Android application. We only consider users for which we have both kinds of data such that we can compute the three types of features for all users analyzed in our dataset.

Call Detail Records (CDRs): The CDRs contain metadata about mobile phone calls (voice) and SMS events. For privacy reasons, no conversational or textual content is recorded. Any personal information available in the metadata (*i.e.* actual phone number) had been previously pseudo-anonymized[5] with an encrypted hash. The metadata available for each pseudo-anonymized phone call record is: *callerID* (encrypted), *calleeID* (encrypted), *type of the call* (incoming/outgoing), *timestamp* of when the call took place, its *duration* – only in the case of voice calls – and the *cell tower ID* through which the phone call was routed. We use the CDRs to compute the Social/Communication features described in Table 1.

Android app: In addition to the CDRs, we analyze data collected through an Android application installed in hundreds of thousands of devices in the United Kingdom. Upon installation, the app –which runs in the background– asks its users for explicit informed consent to record at regular intervals the state of the device, its usage and the context where it is used (*e.g.*, GPS coordinates). To uniformly model application usage, we consider only those apps available in the Google Play Store, which is the main store for Android devices. We use the data collected through this Android app to compute the App Usage and Mobility features described in Table 1.

Note that we consider only data generated by users having GPS locations covering at least 80% of the hours of each individual.

All data is pseudo-anonymized, which means that all personal information is encrypted to preserve privacy. The encryption is consistent for the same *callerID* and *calleeID*, such that we can match the data from the same customers over time and in different datasets. Moreover, all the data had been collected according to existing data protection legislation and all analyses have been carried out following the code of conduct and ethics defined by the institution where the data was collected and analyzed. Finally, all analyses are carried out in aggregated form to preserve privacy.

4.2 Multi-source spatio-temporal data matching

Given that we are dealing with spatio-temporal data, it is important to maximize the spatial and temporal overlap of the data from the different data sources to minimize potential errors due to temporal or spatial mismatches.

From a *spatial* perspective, we aggregate all variables at a *local authority* level. Local authorities are geographically and administratively distinct entities in the UK that have their own local government. There are a total of 418 principal local authorities in the UK.

Values for SWB are available at a local authority level for 389 local authorities in the entire United Kingdom. We drop local authorities for which there is no SWB available in at least one of the two years –which was the case for 3 local authorities: Isles of Scilly, City of London and Richmondshire, leaving 386 local authorities. The sample size of the SWB survey is roughly 165,000 individuals per year, that is roughly 430 respondents on average per local authority.

In contrast to SWB, the IMD is restricted to England.

We use local authority averages for the IMD (326 values as there is only IMD data for England). Hence, sample sizes for both SWB and the IMD are comparably large. Given that the IMD of a local authority is the mean of the ordinal scaled values of the small areas within, the local authority values can be considered as quasi-metric.

Regarding the *temporal* overlap, our mobile data had been captured between January 2017 and August 2018. In order to maximize the overlap with the target variables, we use SWB from 2017 (survey period from April 2016 till March 2017) and 2018 (survey period from April 2017 till March 2018). In terms of the IMD, we use the latest published IMD which was for 2015. As it is also the case in previous work, *e.g.* [34], there is no exact temporal overlap between the IMD and the mobile data. However, the IMD is supposed to be a stable measure, such that using mobile data from 2-3 years after the IMD was collected should still yield valid results.

In terms of the mobile data, there are between 1 million to 1.5 million anonymous users (depending on the year and the number of weeks of analysis) for whom we can assign a local authority as their home location. Hence, on average there is data for 2,600-2,900 users per local authority. To ensure data privacy we define an absolute aggregation threshold of a minimum 50 users per local authority. All local authorities have at least 50 users except for the Isles of Scilly. Therefore we also exclude this local authority for the IMD analysis. In sum, all mobile data is pseudonymized, aggregated to a minimum number of 50 users, in compliance with existing data privacy and protection regulations and analyzed following a strict code of conduct and ethics defined and approved by our organization.

4.3 Feature Extraction

From the mobile data, we compute a set of 16 features aggregated at a weekly level for each of the first three months of 2017 and 2018.

[5] GDPR Article 4(5)

Type	Feature	Description
Social/	Degree [10, 32, 31, 7, 25]	Number of unique contacts with whom the user was in contact, including incoming and outgoing calls and SMS
Comm.	Number of Calls [32, 13, 4, 22]	Number of successful calls, incoming and outgoing
(CDRs)	Number of SMS [32, 13, 22]	Number of SMS, incoming and outgoing
	Aggregated time spent calling [10, 32, 13, 31]	Time calling, incoming and outgoing calls
	Entropy of Contacts [10, 31, 20, 11, 25]	Standardized entropy[a] of the distribution of all communications (incoming and outgoing calls and SMS) over all unique contacts
App Usage	Capacity of Apps [6]	Number of unique apps used
	Number Apps Opened	Number of times the user opened any app
(Android	Aggregated screen time on	Time in which the screen was turned on
app)	Aggregated Data Usage	Up and Downloads (WiFi and Mobile Data)
Mobility	Commuting Distance	Haversine distance between home location and work location. None if no work location
(Android	Distance Traveled [32, 13]	Cumulated Haversine distances for every chronologically consecutive pair of visited stop locations
app)	Radius of Gyration [32, 13, 23, 25]	Average Haversine distance of the stop locations to the center location. The center location is the weighted mean of the coordinates of the stop locations. It is weighted by the time spent at each location
	Capacity of Locations [32, 13, 11, 6]	Number of unique visited stop locations
	Stability of Locations [6]	Number of unique visited stop locations that were also visited in the last week (only to be calculated with at least 2 weeks of data)
	Gain of Locations [6]	Number of unique visited stop locations that were not visited in the last week (only to be calculated with at least 2 weeks of data)
	Entropy of Locations [10, 23, 11, 25]	Standardized entropy[a] over the distribution of summed up time spent at different stop locations

Table 1. Description of the 16 features used in the analysis. Note that all Mobility and App Usage features are derived from the Android app data and all Social/ Communication features are derived from CDR data. All features are aggregated temporally for at least one week and spatially at a local authority level. We include after each feature name citations to related work that used the same features. [a] We use this formula for the Standardized Entropy: $D(i) = \frac{\sum_{j=1}^{k} p_{ij} \log(p_{ij})}{\log(k)}$ where k is the number of contacts/stop locations of user i with $p_{ij} = \frac{V_{ij}}{\sum_{j=1}^{k} V_{ij}}$ where V_{ij} is the volume of calls between user i and user j / the time a user i spent at stop location j.

The features are described in Table 1. As shown on the Table, we use CDRs to compute the Social/Communication features and the Android app to compute the App Usage and Mobility features.

The rationale for defining this battery of features is based on previous work –shown next to each feature on the Table– and on a preliminary feature analysis process where we defined tens of features and selected those with the lowest inter-correlations. In addition to the features found in the literature, we included several new features in context of this task: (1) three features related to mobile App Usage; and (2) Commuting Distance, given that its association with SWB (*e.g.* [19]). Besides features capturing different aspects of quantitative mobility (*e.g.* Distance traveled, Radius of Gyration,. . .), we also calculate features that describe the diversity and dynamics of visited locations (Stability, Gain and Entropy). These features are based on recent work by [6].

In order to analyze the impact of the amount of available data on the quality of the predictions, we compute the means for every user over 1, 2, 4, 8 and 12 weeks. Note that we avoid public or bank holidays to obtain typical, non-seasonal data. As already explained, we aggregate the features spatially at a local authority level given the user's home location.

We use the median rather than the mean values of the features in each local authority given the long-tailed distribution of most of the features. Histograms of these medians show normally distributed values for all 16 features. Thus, the final data matrix consists of 16 features for 326 local authorities for the IMD and for 384 local authorities for SWB computed using each of the possible time windows of training data (1, 2, 4, 8 and 12 weeks).

5 Results

We train different classification and regression supervised machine learning models to automatically infer the 2017 SWB and the 2015 IMD using the features computed from 2017 mobile phone data. Next, analyze the impact of different feature groups on the quality of the predictions for both target variables.

We follow the same procedure in all the supervised machine learning tasks. We train different classification (namely, Logistic Regression, k-Nearest-Neighbours, Linear Support Vector Machines, SVMs with different types of kernels, Random Forests and Gradient Boosted Trees) and regression (Ridge Regression) models as available in the scikit-learn Python package [24].

In terms of the evaluation metrics and given that we have fully balanced classes in all tasks, we use *accuracy* to evaluate the classification models. Accuracy is fairly easy to interpret because it is simply the proportion of correctly classified examples. To better understand the models, their strengths and weaknesses, we also report confusion matrices. Our regression models are evaluated using the *explained variance ratio* R^2. This is the de-facto standard metric in social sciences for regression problems. It is also fairly easy to interpret, because it can maximally become 1 (perfect prediction).

Hyper-parameters are tuned using grid-search –or in the case of a very large hyper-parameter space randomized grid search– in a stratified-5-fold cross-validation setting. When having an optimized set of hyper-parameters, we repeat the training and testing by running the stratified-5-fold cross-validation with 50 different random seeds. The reported evaluation metrics are hence the mean of the 50 means of the 5 validation scores in each iteration.

Task 1: Inferring IMD and SWB from mobile data

We carry out three different tasks with regards to the prediction of the IMD and SWB from mobile data.

Task 1.1: Quartile classification

In this task, we split our well-being datasets into four quartiles for every target variable: 8 facets for the IMD plus the aggregate IMD and 4 facets for SWB. We train a 4-class classifier for each target variable. Performing a 4-quartile classification task enables us to have a balanced dataset with 4 classes, including the top and bottom quartiles, which we consider to be particularly important to policy makers as they contain the local authorities which are in the top and bottom 25% regarding their levels of well-being. We think that quartiles are small enough to allow for targeted policy making and big enough to still deliver interpretable results.

We train different classifiers as previously described. Here, we only report results of linear SVMs as they performed consistently well across all target variables and are fairly easy to interpret. We are interested in studying the impact of the amount of available data on the performance of the classifiers. Hence, we compute all the features using a varying number of weeks of data, namely: 1, 2, 4, 8 and 12 weeks. In our experiments, the classifier's performance increased as the amount of data increased, reaching a plateau at 4 to 8 weeks of data, depending on the target variable. Hence, we present our results using 8 weeks of data.

Using linear SVMs and 8 weeks of data, we obtain the results shown in Table 2, column **4-Quartile Classification**. Note that random guessing would lead to an expected accuracy of 25%. Overall, the IMD facets can be predicted substantially better than SWB. The highest scores are achieved for the Income and Employment facet of the IMD (both $acc = 67\%$), followed by the IMD average score with **66%**[6]. Within SWB, the target variable with the highest accuracy is Life Satisfaction ($acc = 46\%$).

Additionaly, Figure 1 shows the confusion matrices for Life Satisfaction and the IMD average score. Both models perform particularly well for the two extreme quartiles, while performance decreases for the two middle quartiles. Note how the IMD model never confuses the two extreme quartiles.

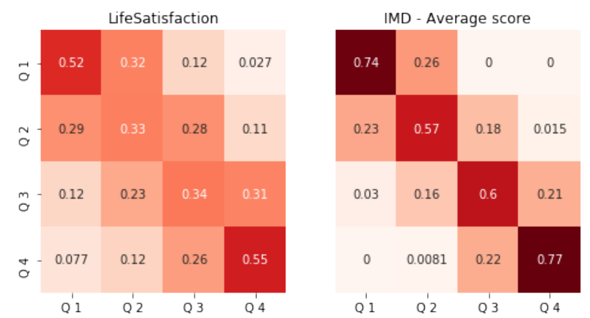

Figure 1. Confusion matrices of the 4-quartile classification task for Life Satisfaction and IMD - Average Score. Shown values are conditional probabilities for a certain prediction (row) under the condition of different actual quartiles (column).

Task 1.2: Binary classification of extreme quartiles

In this next task, we focus on the local authorities that are at the top and bottom quartiles regarding their IMD and SWB. This extreme-quartile binary classification task has been reported previously in the literature [30] and could be valuable to policy makers to inform them where to prioritize their policies and investments. For every target variable, we select the top 25% and bottom 25% local authorities and assign them the values 1 and 0, respectively. Training linear SVMs with 8 weeks of data produced the results shown in the column **Extreme-quartile binary classification** in Table 2.

[6] Note that for the creation of the IMD average score, the Income and Employment facets alone account for 45% of the weighted average index.

While performances improve across all target variables compared to the quartile classification task, the improvement is particularly striking for the average IMD and almost all of its facets. Almost perfect classification (**99%**) was achieved for the average IMD and the income and employment facets.

Task 1.3: Regression

Since both the IMD and SWB values in each local authority are computed by averaging, we can assume them to be continuous.[7] The advantage of formulating a regression problem is that it enables us to differentiate within classes when compared to the classification tasks described above. Furthermore, we are able to compare our results with previous studies that report explained variance ratios or raw correlation (*e.g.* [10]). Using ridge regression, we obtain the explained variances shown in the corresponding column in Table 2. Analogously to the classification tasks, we observe the best performance for the IMD average score ($R^2 = 0.76$). Regarding SWB, the best performance corresponds to the Life Satisfaction facet ($R^2 = 0.29$).

Through these three tasks, we conclude that it is indeed possible to infer aggregate objective and subjective well-being from mobile data. Our classification and regression models yield significantly higher performance when predicting objective well-being (IMD) than subjective well-being (SWB).

We discuss these results and their implications in Section 6.

Group	Target	4-quartile	Extreme-quartile binary	Reg. R^2
IMD 2015	IMD -Average	66%	**99%**	0.76
	Income	**67%**	**99%**	**0.80**
	Employment	**67%**	**99%**	**0.80**
	Education, Skills and Training	54%	96%	0.63
	Health Depriv. and Disability	65%	98%	0.75
	Crime	61%	98%	0.71
	Barriers to Housing and Services	49%	93%	0.48
	Living Environ.	49%	87%	0.42
SWB 2017	Life Satisfaction	**46%**	**84%**	**0.29**
	Worthwhile	40%	77%	0.20
	Happiness	38%	80%	0.19
	Anxiety	37%	72%	0.15

Table 2. Results for the classification and regression tasks using 8 weeks of training data from 2017. Reported values are accuracy for classification (columns 3 and 4) and explained variance ratio for regression (column 5). Note that a random classifier would correspond to 25% accuracy and 50% accuracy for the 4-quartile and binary classification problems, respectively

Task 2: Feature importance analysis

Next, we study how the three different groups of features (Social/Communication, App Usage and Mobility) are suited to predict our target variables. Given the previously reported results, we carry out the feature importance analysis on two target variables: the average IMD score and Life Satisfaction within SWB. We focus on Life Satisfaction because Life Satisfaction is the only evaluative facet in the SWB survey and therefore the most stable measure of subjective well-being of the four available facets.

We report results of our feature importance analysis for the 4-quartile classification problem. Note that we obtained similar feature importance results in the other tasks (binary classification and regression) and hence do not report them due to space constraints.

[7] The IMD values are the means of the small areas' IMD values within a local authority and the SWB values are the means of the respondents' answers aggregated by their home local authority.

Even though we selected the features to have minimal inter-correlations, in a preliminary analysis we identified inter-correlations between some of the features. It is known that high co-linearity of features leads to deceptive interpretations of feature importances[8]. Thus, we did not use the SVM coefficients in our feature importance analysis. Instead, we assess the importance of the three different feature groups by training models which take as input all possible combinations of the three feature groups. As a result, we compute 7 models for each target variable (the average IMD score and Life Satisfaction). The results are shown in Figure 2.

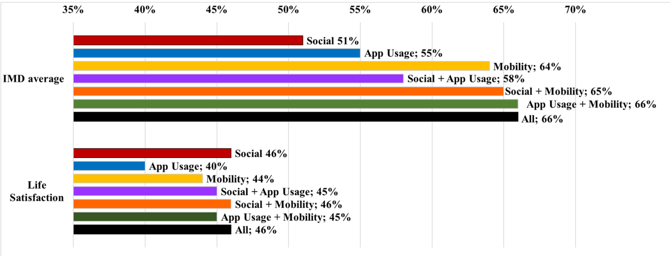

Figure 2. Accuracies of a 4-quartile classification problem with different subsets of features

As seen in the Figure, the model that uses all the features delivers the best predictions for the average IMD (acc = 66%). The model trained only with Mobility features yields a competitive accuracy of 64%. Adding App Usage features to the Mobility features increases the accuracy to that of the full model (acc = 66%).

Regarding Life Satisfaction, the classifier trained with Social/Communication features alone yields the best performance (acc = 46%). The second most predictive group of features is Mobility (acc = 44%), followed by App Usage (acc = 40%) features.

To further investigate the role of individual features, we calculate the Pearson correlation between individual features and the two target variables. A visual inspection of the scatter plots revealed that all associations are best described by linear correlations. Therefore we did not carry out any data transformation. Table 3 depicts the correlations between the individual features and our target variables (the average IMD and Life Satisfaction).

As seen in the Table, we find strong negative correlations between the average IMD and Mobility features: Distance traveled ($r = -0.77$) and Radius of Gyration ($r = -0.68$), and a strong positive correlation with one App Usage feature: App Usage ($r = 0.68$) and one Social/Communication feature: Degree ($r = 0.59$).

In terms of Life Satisfaction, we find strong positive correlations with two Mobility features: Distance traveled ($r = 0.51$) and Radius of Gyration ($r = 0.54$) and strong negative correlations with Social/Communication features: Number of Calls ($r = -0.49$) and Time Calling ($r = -0.50$).

Hence, the more people move, the lower their Index of Multiple Deprivation and higher their Life Satisfaction. These results are aligned with previous work which has found that the larger the mobility, the higher the socio-economic status of a region [32]. The positive correlation between Commuting Distance and Life Satisfaction might seem to contradict previous studies (*e.g.* [5]). Moreover, the more time people spend calling and the highest their data usage of their smartphones, the higher their Index of Deprivation and the lower their Life Satisfaction. We discuss these findings next.

[8] For a demonstration see https://explained.ai/rf-importance/index.htmlcollinear

6 Discussion and Implications

Given the results obtained, we discuss here our main findings, the relationship between our findings and previous work and outline a few implications for the design of ubiquitous mobile systems to infer aggregate objective and subjective well-being.

Finding 1: Aggregate IMD can be inferred more accurately than SWB from passively collected mobile phone data

We obtained **66%** accuracy for the average IMD and **46%** accuracy for Life Satisfaction (SWB facet) in a 4-quartile classification task. Distinguishing the top and the bottom quartiles yields almost perfect results for the average IMD (**99%** accuracy) and competitive results for Life Satisfaction (**84%** accuracy). Continuous regression models deliver $R^2 = 0.76$ for the IMD and $R^2 = 0.29$ for Life Satisfaction. The almost perfect accuracy (99%) in classifying local authorities in the top vs. the bottom quartiles of IMD supports the use of mobile phone data for decision making and leads to our first implication for the design: *our methodology could be used to automatically identify the most deprived regions and design public policies that prioritize such regions*. The best predicted facets of the IMD are employment and income with 99% accuracy and $R^2 = 0.80$.

Regarding aggregate SWB, the first observation is that it is harder to infer from mobile phone data than objective well-being. Within SWB, the models deliver the best performance on the Life Satisfaction facet as shown in Table 2. Life Satisfaction falls within the evaluative part of well-being, which is rather stable and therefore does not need to be measured at the exact same point in time as the behavioral data was collected. Conversely, Happiness and Anxiety –which are examples of the affective dimension of well-being –are more fragile constructs [9] and hence harder to model in aggregate form.

A second implication for the design is that *mobile phone data is valuable to infer aggregate objective well-being*. However, subjective well-being is a more complex construct which seems to require different data sources or methods. In terms of subjective well-being, our third implication is: *Life Satisfaction is the facet of aggregate subjective well-being most correlated with human behavioral data passively captured by mobile phones.* At the same time, the loss of variance in SWB due to aggregation leads to a fourth implication: *our findings might raise questions regarding the value of having aggregated SWB in regions such as local authorities in the UK.*

Finding 2: Mobility and Social/Communication features are key to infer aggregate IMD and SWB, respectively

Different types of features –reflective of different aspects of human behavior– are the most predictive to infer objective vs. subjective well-being. Figure 2 and Table 3 summarize our feature importance analyses. Figure 2 shows the performance of a battery of models which use all possible combinations of the three different types of features (Social/Communication, App Usage and Mobility) to perform a 4-quartile classification task on IMD and the Life Satisfaction facet of SWB.

Our findings are aligned with the literature. First, we find that Mobility features are highly predictive of the average IMD, which is consistent with previous work, e.g. [10].

Second, Social/Communication features are the most useful to infer Life Satisfaction. While we have not found any previous work aimed at automatically inferring aggregate Life Satisfaction from large-scale passively collected human behavioral data, this finding is consistent with previous work on a smaller scale [2, 28].

Next, we focus on the correlations between individual features and our two target variables (IMD and Life Satisfaction), shown in Table 3.

Regarding the IMD, we observe that Call Volume (Number of phone calls and Time spent in phone calls) and the Degree of the call graph are positively correlated with the IMD, *i.e.* the more in-

Group	Feature	SWB Life Sat	IMD Avg.
Social Comm.	Degree	-0.46	0.59
	Number Calls	-0.49	0.57
	Number of SMS	-0.13	0.40
	Time Calling	-0.50	0.48
	Entropy of Contacts	-0.22	0.00
App Usage	Capacity of Apps	-0.23	0.08
	Apps Opened	-0.24	0.22
	Screen Time	-0.21	0.34
	Data Usage	-0.44	0.68
Mobility	Commuting Distance	0.37	-0.56
	Distance Traveled	0.51	-0.77
	Radius of Gyration	0.54	-0.68
	Capacity Locations	-0.19	-0.21
	Stability Location	0.10	0.32
	Gain Locations	-0.25	0.04
	Entropy Location	-0.09	-0.22

Table 3. Correlation between individual features and the two target variables (Life Satisfaction and average IMD).

tense the use of the phone and the larger the number of contacts, the higher the index of deprivation. Conversely, most Mobility features are strongly (Commuting Distance, Distance Traveled and Radius of Gyration) to moderately (Capacity of Locations and Entropy of Locations) negatively correlated with the IMD. These correlations with objective well-being have also been partially found in previous work [10] which was also carried out in the UK.

Also in accordance with our results, [32] find strong positive associations between distance-related Mobility features (distance traveled and radius of gyration) and the socio-economic level (inverse of deprivation) of a region. Finally, App and Data Usage (Screen Time, Apps Opened and Data Usage) are strongly positively correlated with the IMD. While we do not have conclusive evidence to explain this finding, we believe that it might reflect the amount of time people spend at home consuming large volumes of data, such as watching videos or playing video games. People living in areas with high IMD tend to have lower Radius of Gyration (as previously explained) and higher unemployment and hence spend more time at home, which might explain the positive correlation between IMD with Data Usage.

With regards to SWB, Call Volume (Number of calls, Time calling) and the Degree are negatively correlated with Life Satisfaction. In other words, the more time people spend calling on their mobile phones and the larger the number of contacts, the lower their Life Satisfaction, which might seem surprising. As reported in the literature [18, 2], *good* and frequent social relations positively affect Life Satisfaction. However, intensity of usage of the mobile phone and the size of a person's contact list does not necessarily imply high quality, meaningful relationships.

On the other hand, Mobility features (Distance Traveled, Radius of Gyration and Commuting Distance) are positively correlated with Life Satisfaction. This positive correlation with commuting distance is particularly puzzling given previous work which has found that commuting has a negative impact on Life Satisfaction (see *e.g.* [14]). However, our research is based on quantitative and passively collected human behavioral data – as opposed to self-reported behavior – and there is recent work that also supports our findings [19]. It seems that additional factors such as the quality of the commute need to be taken into account. Finally, we find a negative correlation between Apps and Data Usage and Life Satisfaction. We hypothesize that Data Usage might be reflective of unemployment which is known to have a negative relationship on Life Satisfaction [36]. We find a positive correlation of 0.60 between Data Usage and the unemployment facet of the IMD. We leave to future work a further exploration of these findings.

Our fifth implication derives from our main finding in this section: *Mobility features might be sufficient to infer aggregate objective well-being and Social/Communication features to infer aggregate Life Satisfaction.*

Finding 3. Results are not universally applicable

Our findings are certainly not universally applicable.

While we observe a positive correlation between intensity of usage of the mobile phone and deprivation, the correlation between these features and socio-economic status is reversed in developing economies, *e.g.* [31].

This difference might be due to the fact that in developed countries –like the one under study, people use a wider variety of communication tools beyond traditional phone calls –such as WhatsApp or Facebook Messenger– even for voice calls. Conversely, poorer regions where smartphones are less pervasive and affordable, GSM-based calls and SMS are still prevalent and an indicator of economic prosperity.

Thus, our sixth implication would be *to avoid generalizations of the results to other geographies or even to other moments in time*.

7 Conclusions and Future Work

In this paper, we have described MobiSenseUs, a machine-learning system to automatically infer aggregate objective and subjective well-being in the UK from two types of aggregate and pseudonymized passively collected mobile data. MobiSenseUs is framed within a global movement of exploring novel data sources to assist national statistics offices and governments in achieving more efficient and evidence-based policy making.

Our work has contributed to the state-of-the-art in several ways. First, we have analyzed one of the largest to date pseudonymized sample of rich human behavioral data as captured by two different mobile phone data sources over several months. We have computed human behavioral features regarding Communication/Social Connections, App Usage and Mobility. Second, this is the first work to report results for *both* spatially aggregated objective and subjective well-being in the same large-scale quantitative study. Third, we have built several machine learning models to automatically classify local authorities according to their well-being. Our results demonstrate the potential of mobile phone data for the efficient appropriation of public funds. We have found that IMD is easier to predict and model from aggregate mobile data than SWB. Fourth, we have performed a feature importance analysis to shed light on the role that different types of human behavioral features play on well-being. Mobility features are particularly important to predict IMD whereas Social/Communication features are the most predictive features for SWB (Life Satisfaction in particular).

While our results are promising and interesting, they uncover several important areas of future work, including: (1) a more in-depth study of the negative correlations between Life Satisfaction and the intensity of usage of the phone and the degree of the call graph; and (2) of the positive correlations between commuting distance and Life Satisfaction; (3) a longitudinal study over a longer period of time to shed light on the temporal dependencies of our findings; (4) a more thorough investigation of the relationship between App and Data Usage and objective and subjective well-being; and (5) a replication of our study in a different country to better understand the impact of geography and culture on our results.

REFERENCES

[1] Tim Althoff, Jennifer L Hicks, Abby C King, Scott L Delp, Jure Leskovec, et al., 'Large-scale physical activity data reveal worldwide activity inequality', *Nature*, **547**(7663), 336, (2017).

[2] Viviana Amati, Silvia Meggiolaro, Giulia Rivellini, and Susanna Zaccarin, 'Social relations and life satisfaction: the role of friends', *Genus*, **74**, (12 2018).

[3] Joshua Blumenstock, Gabriel Cadamuro, and Robert On, 'Predicting poverty and wealth from mobile phone metadata', *Science*, **350**(6264), 1073–1076, (2015).

[4] Galo Castillo, Fabricio Layedra, Maria-Belen Guaranda, Paolo Lara, and Carmen Vaca, 'The silence of the cantons: Estimating villages socioeconomic status through mobile phones data', in *2018 International Conference on eDemocracy & eGovernment (ICEDEG)*, pp. 172–178. IEEE, (2018).

[5] Kiron Chatterjee, B. Clark, A. Martin, and A. Davis, 'The commuting and wellbeing study: Understanding the impact of commuting on people's lives', Technical report, UWE Bristol, UK.

[6] Marco de Nadai, Angelo Cardoso, Antonio Lima, Bruno Lepri, and Nuria Oliver, 'Strategies and limitations in app usage and human mobility', *Nature Scientific Reports*, **9**(10935), (2019).

[7] Adeline Decuyper, Alex Rutherford, Amit Wadhwa, Jean-Martin Bauer, Gautier Krings, Thoralf Gutierrez, Vincent D Blondel, and Miguel A Luengo-Oroz, 'Estimating food consumption and poverty indices with mobile phone data', *arXiv preprint arXiv:1412.2595*, (2014).

[8] Ed Diener, 'Subjective well-being.', *Psychological bulletin*, **95**(3), 542, (1984).

[9] Paul Dolan, Richard Layard, and Robert Metcalfe, 'Measuring Subjective Well-being for Public Policy ', Technical report, Office for National Statistics, (2011).

[10] Nathan Eagle, Michael Macy, and Rob Claxton, 'Network diversity and economic development', *Science*, **328**(5981), 1029–1031, (2010).

[11] Gregor Engelmann, Gavin Smith, and James Goulding, 'The unbanked and poverty: Predicting area-level socio-economic vulnerability from m-money transactions', *2018 IEEE International Conference on Big Data (Big Data)*, 1357–1366, (2018).

[12] Vanessa Frias-Martinez, Cristina Soguero-Ruiz, Enrique Frias-Martinez, and Malvina Josephidou, 'Forecasting socioeconomic trends with cell phone records', in *Proceedings of the 3rd ACM Symposium on Computing for Development*, p. 15. ACM, (2013).

[13] Vanessa Frias-Martinez and Jesus Virseda, 'On the relationship between socio-economic factors and cell phone usage', in *Proceedings of the fifth international conference on information and communication technologies and development*, pp. 76–84. ACM, (2012).

[14] M. Hilbrecht, B. Smale, and S.E. Mock, 'Highway to health? commute time and wellbeing among canadian adults', *World Leisure Journal*, (56 (2)), 151–163, (2014).

[15] Desislava Hristova, Matthew J Williams, Mirco Musolesi, Pietro Panzarasa, and Cecilia Mascolo, 'Measuring urban social diversity using interconnected geo-social networks', in *Proceedings of the 25th international conference on world wide web*, pp. 21–30. International World Wide Web Conferences Steering Committee, (2016).

[16] Neal Lathia and Licia Capra, 'How smart is your smartcard?: measuring travel behaviours, perceptions, and incentives', in *Proceedings of the 13th international conference on Ubiquitous computing*, pp. 291–300. ACM, (2011).

[17] Neal Lathia, Daniele Quercia, and Jon Crowcroft, 'The hidden image of the city: sensing community well-being from urban mobility', in *International conference on pervasive computing*, pp. 91–98. Springer, (2012).

[18] J. Lee and J. Cagle, 'Social exclusion factors influencing life satisfaction among older adults', *Journal of Poverty and Social Justice*, (26 (1)), 35–50, (2018).

[19] Olga Lorenz, 'Does commuting matter to subjective well-being?', *Journal of Transport Geography*, **66**, 180–199, (January 2018).

[20] Huina Mao, Xin Shuai, Yong-Yeol Ahn, and Johan Bollen, 'Mobile communications reveal the regional economy in côte d'ivoire'.

[21] Afra Mashhadi, Sourav Bhattacharya, and Fahim Kawsar, 'Understanding the impact of geographical context on subjective well-being of urban citizens', in *Proceedings of the Second International Conference on IoT in Urban Space*, pp. 29–35. ACM, (2016).

[22] Christopher Njuguna and Patrick McSharry, 'Constructing spatiotemporal poverty indices from big data', *Journal of Business Research*, **70**, 318–327, (2017).

[23] Luca Pappalardo, Dino Pedreschi, Zbigniew Smoreda, and Fosca Giannotti, 'Using big data to study the link between human mobility and socio-economic development', in *2015 IEEE International Conference on Big Data (Big Data)*, pp. 871–878. IEEE, (2015).

[24] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, 'Scikit-learn: Machine learning in Python', *Journal of Machine Learning Research*, **12**, 2825–2830, (2011).

[25] Neeti Pokhriyal and Damien Christophe Jacques, 'Combining disparate data sources for improved poverty prediction and mapping', *Proceedings of the National Academy of Sciences*, **114**(46), E9783–E9792, (2017).

[26] Daniele Quercia, Jonathan Ellis, Licia Capra, and Jon Crowcroft, 'Tracking gross community happiness from tweets', in *Proceedings of the ACM 2012 conference on computer supported cooperative work*, pp. 965–968. ACM, (2012).

[27] Daniele Quercia and Diego Saez, 'Mining urban deprivation from foursquare: Implicit crowdsourcing of city land use', *IEEE Pervasive Computing*, **13**(2), 30–36, (2014).

[28] Michael A. Shields, Stephen Wheatley Price, and Mark Wooden, 'Life satisfaction and the economic and social characteristics of neighbourhoods', *Journal of Population Economics*, **22**(2), 421–443, (Apr 2009).

[29] Tom Smith, Michael Noble, Stefan Noble, Gemma Wright, David McLennan, and Emma Plunkett, 'The english indices of deprivation 2015', *London: Department for Communities and Local Government*, (2015).

[30] Chris Smith-Clarke, Daniele Quercia, and Licia Capra, 'Finger on the pulse: identifying deprivation using transit flow analysis', in *CSCW*, (2013).

[31] Christopher Smith-Clarke, Afra Mashhadi, and Licia Capra, 'Poverty on the cheap: Estimating poverty maps using aggregated mobile communication networks', in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 511–520. ACM, (2014).

[32] Victor Soto, Vanessa Frias-Martinez, Jesus Virseda, and Enrique Frias-Martinez, 'Prediction of socioeconomic levels using cell phone records', in *International Conference on User Modeling, Adaptation, and Personalization*, pp. 377–388. Springer, (2011).

[33] Jessica E Steele, Pål Roe Sundsøy, Carla Pezzulo, Victor A Alegana, Tomas J Bird, Joshua Blumenstock, Johannes Bjelland, Kenth Engø-Monsen, Yves-Alexandre de Montjoye, Asif M Iqbal, et al., 'Mapping poverty using mobile phone and satellite data', *Journal of The Royal Society Interface*, **14**(127), 20160690, (2017).

[34] Alessandro Venerandi, Giovanni Quattrone, Licia Capra, Daniele Quercia, and Diego Saez-Trumper, 'Measuring urban deprivation from user generated content', in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, CSCW '15, pp. 254–264, New York, NY, USA, (2015). ACM.

[35] Alan S Waterman, 'Two conceptions of happiness: Contrasts of personal expressiveness (eudaimonia) and hedonic enjoyment.', *Journal of personality and social psychology*, **64**(4), 678, (1993).

[36] M. Wulfgramm, 'Life satisfaction effects of unemployment in europe: The moderating influence of labour market policy', *Journal of European Social Policy*, (24 (3)), 258–272, (2014).