# Evaluating the performance of a linear forecasting model and a standard ML method for mood prediction

B. Manzato[1] [2737818], P. Soteriou[1] [2738068], and N. Tabarini[1] [2748673]
Data Mining Techniques - Group 117

MSc Bioinformatics and System Biology - Faculty of Science, Vrije Universiteit (VU) Amsterdam, NL

## 1 Introduction

Extensive smartphone use contributes to low mood and other mental health detriments to broad populations, including teenagers, young and older adults, regardless of geographic location [5, 1, 25, 26, 11, 17]. The ownership and use of smartphones and wearable technology, such as activity trackers, has been experiencing a steady increase and is expected to increase even more over the coming years [14, 24]. Surveys have shown that a large number of people routinely spend a significant amount of time,  4 hours, on their phones daily [6]. Increased and widespread smartphone use and higher rates of mental health disorders have been temporally correlated, indicating that mental health disorders will likely continue to climb as a consequence of increased smartphone use[25].

Increased use and, thus, sampling of smartphone use and wearable technology data signifies that there will be a massive expansion in personal health and behavioural data collected from these devices. Mining these has the potential to uncover a range of insights about the negative effect of lifestyle and environment on physical and mental health. Furthermore, using these data can identify less problematic ways of using smartphones.

Conversely, smartphone and wearable device data mining can shed light to many other positive physiological processes and advance biological research. Namely, the field of circadian rhythms would be provided with rich and novel data on how different behaviours and activities are affected by environmental and ecological factors, seasonal change and consequent daytime duration change. Potentially, even new obscure circannual human behaviours could be discovered [7]. Collection of data of a single individual's behaviours and patterns could identify and predict upcoming depressive episodes, provide insights on their prevention, thereby shortening them via intervention.

Mining of smartphone and activity data that was collected over time, i.e. time-series or temporal data, is challenging and additional processing or more complex methods of analysis are required to model and make predictions based on them. Such approaches include the sliding window protocol and autoregressive integrated moving average (ARIMA) which can handle the sequential nature of the data. The sliding window protocol considers a smaller, fixed-size part of the dataset, which is iteratively and in a stepwise manner slid over the whole dataset considering new data points at each implementation. ARIMA essentially performs linear regression and considers its own lags and lagged forecast errors as predictors. Its implementation requires the time series to be stationary by differencing the values, if needed. The number of differencing steps, d, to achieve a stationary time series is a hyperparameter for ARIMA. Furthermore, the model requires the order of the auto-regressive term, p, which refers to the number of lags to be used as predictors, and the order of the moving average term, q, which represents the number of lagged forecasting errors. Overall, the ARIMA model is the sum of a linear combination of its lags, a linear combination of lagged forecast errors and a constant, as depicted in the following equation:

$$Y_t = \alpha + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + .. + \beta_p Y_{t-p} \epsilon_t + \phi_1 \epsilon_{t-1} + \phi_2 \epsilon_{t-2} + .. + \phi_q \epsilon_{t-q} \quad (1)$$

In this experiment we investigated whether we can predict mood using data collected from smartphone data, using standard machine learning (ML) algorithms, including support vector regression (SVR), and a more sophisticated ARIMA based model. The 19 variables measured were used to identify the most important features for mood prediction and the dataset was used to train the two models. Their performances were then compared to a basic baseline predictor to identify the best performing model, able to predict the mood of a subsequent, unlabeled day with higher accuracy.

## 2 Methods

### 2.1 Dataset description

We used a dataset consisting of data collected using an unobtrusive ecological momentary assessment (EMA) application from the mobile phones of 27 healthy subjects. A total of 19 variables were measured, three of which were self-reported: mood, logged at 5 set time points during the day, circumplex arousal (representing alertness) and circumplex valence (representing a pleasant-displeasant continuum). Time and/or duration of call and text message events, screen on and off time and app usage were also recorded. The apps were then further classified into types: built-in, communication, entertainment, other, social, unknown, utilities, finance, office, travel, weather and game. Lastly, data collected from an accelerometer were used as a proxy for activity.

The dataset contained data for a mean number of days of 42.9 per subject. The dates of each subject's data collection are not the same for all of them, however there is some overlap in some cases.

### 2.2 Dataset preprocessing

To bring the dataset to a more workable/intuitive format ('preprocessed dataset'), the dataset was ordered based on patients id and date. Consequently, the time-of-day attribute of the date was disregarded, and for all other features the mean or sum per single day was used. An average was calculated for mood, activity, valence and arousal. Sum was computed for all the variables measuring time spent on a specific app and for sms and calls. Rows that had values for just call and/or sms were removed, because they were deemed uninformative. The prediction can be considered reliable only if it used measurements from the days directly previous to the day we want to predict, for this reason non-consecutive days were removed. Any NAs remaining were assumed to be not registered and then transformed to zeros for variables that measure time. For the mood variable, NAs were transformed to the mean mood value per subject.

### 2.3 Support Vector Regression

**Preprocessing for SVR** A windowing approach was followed: we averaged the data of the preprocessed dataset over a window of size n, as we thought it was important to take more than one previous day into consideration as it is common practice with temporal data [3, 18, 21]. The windowing process consisted in taking a window of 5 consecutive days; 1 to 5 were used to predict the mood of day 6. The windows were created in a stepwise manner (i.e. the next window is days 2 to 6, to predict day 7).

An extra variable was added, called target mood. First, the NAs for the target variable were filled in with the mean of the mood per patient. With the windowing approach a prediction cannot be created before the 6th day. For this reason we shifted the target variable up by 5 days for each subject.

These preprocessing steps allowed us to break the temporal relation between predictor variables and target variable and treat the data as a common dataset for which a prediction with a standard ML approach can be made.

The problem was approached in a non-personalised way: a single model was created from the data from all patients together. The data was split in training and tests in the proportion of 70% and 30%; the training and test sets consist of 872 and 288 observations respectively.

**SVR** The goal of support vector machine (SVM) is to discover a feature-separating hyperplane in an n-dimensional space, where n denotes the number of features or independent variables. SVM can also be used for regression problems with SVR. SVR maintains all the main features that characterise the algorithm with the goal to minimise the error observed during the training process to get a generalised performance of continuous values [4].

The model was built using the *SVR* method from the *sklearn* package.

**Feature Selection** Feature selection in ML can be defined as the task of selecting an optimal subset of features for a target outcome. The main advantage is that removing features reduces the dataset and avoids the risk of overfitting.

R package *caret* was used to perform recursive feature elimination (RFE). RFE starts with all features in the training dataset, fitting the given ML algorithm (linear regression in this case), ranking features by importance, and removes the least important features, then re-fits the model until the desired number remains. From the original 19 features, RFE found an optimal subset of 14 variables. This number was consistent with the output from another feature selection method (from R package *boruta*) that finds all features which are either strongly or weakly relevant to the decision variable.

Figure 1 shows the importance of the considered features:

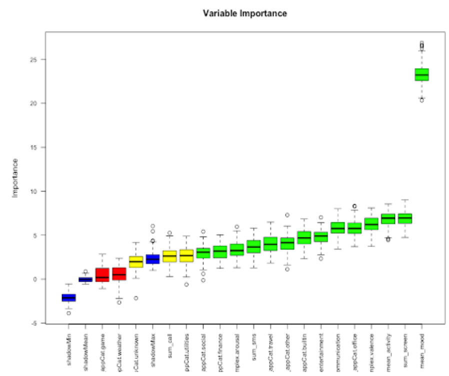


**Fig. 1:** *Importance of each of 19 features, measured by unobstructive ecological momentary assessment obtained using the boruta R package.*

**Hyperparameter optimization** Hyperparameters guide the learning process for a specific dataset, therefore their choice can have a significant impact on the model performance and even small changes can have large effects. Therefore, hyperparameters need to be selected carefully as the model may give better performance with a better selection [2].

A randomized search was implemented to compare the model's performance with different hyperparameter settings. A bounded domain of hyperparameter values was defined and the function RandomizedSearchCV randomly sampled points in that domain to find the best combination. The optimized values were the cost value C and gamma. The best combination turned out to be a radial basis function (RBF) kernel with gamma=0.001 and C=100.

### 2.4 ARIMA

**Preprocessing for ARIMA** To train ARIMA, the date and mean mood values from the preprocessed dataset were used. In the case of overlapping dates, the median mood value was chosen.

**Hyperparameter selection** Optimal model parameters, p, d, q, were identified using auto_arima taken from the *pmdarima* package. The obtained parameters, namely p=0,d=0 and q=1 were selected for building the ARIMA model.

Auto-ARIMA detects optimal parameters by performing differencing tests and subsequently fitting multiple models, aiming to minimise the chosen information criterion. For this analysis the Akaike information criterion (AIC) was chosen.

**Time series cross-validation** The dataset was split into training and test sets, consisting of 70% and 30% of the data, using temporal_train_test_split taken from the *sktime* package. The training set was used to build the model using the

ARIMA method from the *statsmodels* package and the prediction was obtained using forecast.

As the aim is the prediction of the mood for the subsequent day, the time series cross-validation approach was used. Such method consists in initially training the model on the newly-created training set and using it to predict the first instance of the test set. As the prediction was obtained, a new element (originally in the test set) was added to the training set and a new model was fitted [10]. In fact, since a prediction relies on varying weights of the days prior to the prediction, with the closest day weighed more heavily, the day prior to the prediction is required to make predictions [20]. Using this approach, all elements of the selected test set were predicted and the average performance of all the trained models evaluated.

### 2.5 Benchmark Model

The baseline model was implemented differently for the two approaches, as different datasets were used for the prediction.

To evaluate SVR performance, the benchmark was built by simply shifting down the mood value by one in the already processed dataset for windowing. The first instance per subject was then removed, as the prediction was not available.

To assess ARIMA's performance, a Naïve forecaster was implemented using the *sktime* package, specifying the strategy to use as predicting the next day as the previous one (strategy=last).

### 2.6 Metrics for Model Evaluation

The root mean squared error (RMSE) and mean absolute error (MAE) metrics were used to quantify the prediction performance and distance to actual values. Two error metrics were chosen due to varying robustness to outliers, hence providing additional confidence in our results, and due to their common use in regression model evaluation[10].

RMSE is the standard deviation of the residuals, expressing how spread out such values are from the line denoting the best fit. Consequently, smaller RMSE values indicate higher performance.

MAE is defined as the average of the absolute difference between predicted and true values [12]. It is a linear score, meaning that all the individual differences are weighted equally in the average, allowing the metric to be more robust to outliers.

## 3 Results

### 3.1 Exploratory Data Analysis (EDA)

Figure 2A shows the Pearson correlations between features of the preprocessed dataset. Some notable correlations occur, which could be predictive for future mood values, (target). Namely, correlations between the target and mean_mood variables ($r = 0.51$), target and mean_circumplex.valence ($r = 0.25$) give the highest correlation coefficients. Moreover, the correlation between mean_mood and

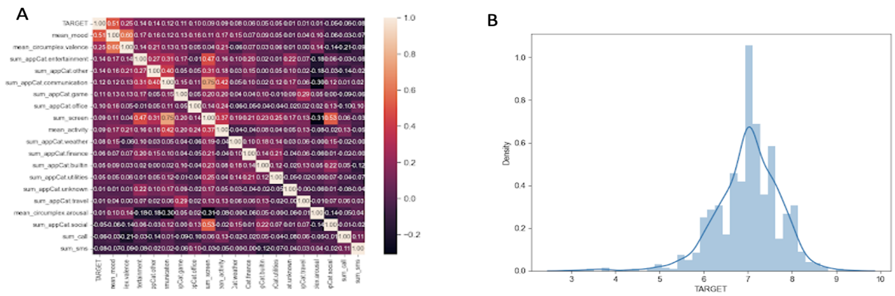


**Fig. 2:** ***A****) Correlation matrix for 19 variables measured using unobtrusive ecological momentary assessment after daily averaging. Colours and numbers indicate Pearson correlation coefficient, r.* ***B****) Histogram of the mean_mood variable values after daily averaging.*

mean_circumplex.valence (r = 0.6) also provides further support for its predictive value for target, since target is essentially a future mood value. Expectedly, the sum_screen variable correlates with: sum_appCat.entertainment (r = 0.47), sum_appCat.communication (r = 0.75) and sum_appCat.social (r = 0.53). Interestingly, mean_circumplex.arousal gives several notable negative correlations with sum_screen (r = -0.31), sum_appCat.communication (r = -0.3) and sum_appCat.entertainment (r=-0.18). Unexpectedly, mean_activity also correlates with sum_appCat.communication (r = 0.42) and sum_screen (r = 0.37).

Figure 2B target density distribution shows the frequency of the occurrence of target mood values averaged within a day. The distribution shows a slight negative skew with a longer tail towards lower values. There is a clear peak at target value 7, and the majority of values occur between values 5.5 and 8.5.

### 3.2 SVR

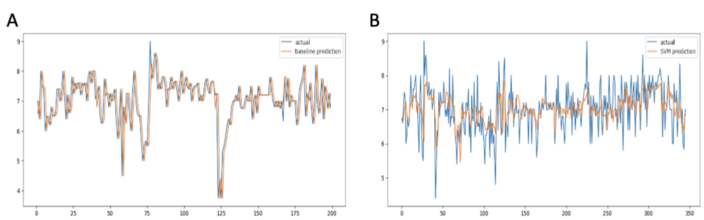


**Table 1:** Errors for SVR and BASELINE models

| | SVR | BASELINE |
|---|---|---|
| RMSE | 0.631 | 0.762 |
| MAE | 0.455 | 0.525 |

**Fig. 3:** ***Baseline and SVR predictions comparison.*** *On the x-axis the number of instances are shown. On the y-axis, the mood values.* ***A****) Baseline prediction (orange) against real values (blue) for the the test set. Just the first 200 instances were plotted to allow better visualization of the 1-day shift.* ***B****) SVR prediction (orange) against real values (blue) for the the test set.*

Figure 3 demonstrates the predicted values for the trained SVR and baseline models and the actual mood values. SVR model predictions capture the overall trend and fluctuations of the real values in the dataset, however predicted values are dampened and not capturing the extreme actual values. SVR only predicts a few values under 6 and over 7.5, whereas there are multiple occurrences of values higher than 7.5 and lower than 6 in the actual values. The baseline model predictions follow exactly the same pattern as the actual values, shifted by 1 day.

The SVR model outperforms the baseline [Table 1]. RMSE and MAE for the SVR model are lower than the baseline, indicating more precise predictions more closely depicting the actual, true values. The difference between the metrics shows an improvement of 13-17% for SVR model predictions.

### 3.3 ARIMA

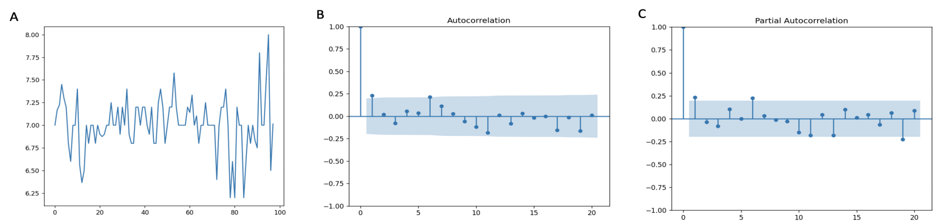


**Fig. 4:** ***Plots for ARIMA hyperparameter selection. A****) Mood value per day after ARIMA preprocessing.* ***B****) Autocorrelation plot for the ARIMA model obtained using the auto_arima package.* ***C****) Partial autocorrelation plot for the ARIMA model obtained using the auto_arima package. 95% confidence intervals shown as blue boxes.*

**ARIMA hyperparameter tuning.** The trend of the data was checked to see if stationary (i.e. there is not a clear trend in time), as ARIMA requires stationary data in order to be applied. The figure shows no particular trend, therefore d = 0 was deemed an appropriate choice [Figure 4A]. Looking at the autocorrelation plot, a single spike and sharp drop is observed, indicating that parameter q should be 1 [Figure 4B]. The partial autocorrelation plot, lastly, shows that no points greatly exceed the 95% confidence interval, indicating that a p parameter value of 0 is optimal, as suggested by auto-ARIMA [Figure 4C].

**ARIMA prediction.** Figure 5 demonstrates the predicted values for the ARIMA and baseline models and the actual mood values. Mood values predicted using the ARIMA model follow the overall trend and fluctuations of the real values, generally with a slight delay of 1 day. However, the actual values were not correctly predicted, indicated by the model not capturing the extreme values on days 10, 12, 16, 23 and 27. The baseline model follows exactly the same pattern as the actual values, shifted by one day.

The ARIMA model outperforms the baseline, based on the error metrics chosen [Table 2]. RMSE and MAE for the ARIMA model are lower than the

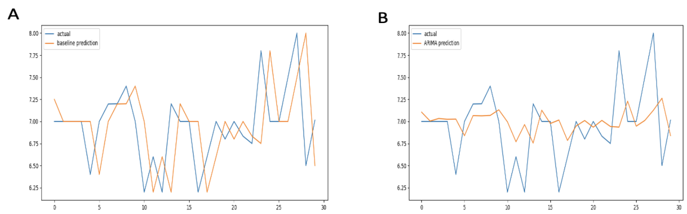


**Table 2:** Errors for ARIMA and BASELINE models

| | ARIMA | BASELINE |
|---|---|---|
| RMSE | 0.420 | 0.543 |
| MAE | 0.305 | 0.405 |

**Fig. 5:** ***Baseline and ARIMA predictions comparison.*** *On the x-axis the number of instances are shown. On the y-axis, the mood values.* ***A****) Baseline prediction (orange) against real values (blue) for the the test set.* ***B****) ARIMA prediction (orange) against real values (blue) for the the test set.*

baseline, indicating that the model predictions more precisely depict the actual, true values. Overall, the difference between the metrics shows a 22%-25% improvement in prediction of the trained ARIMA model over the baseline.

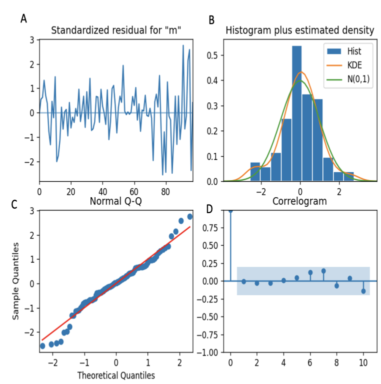


**Fig. 6:** ***ARIMA fit residual error description plots.*** ***A****) Standardised residual errors per day.* ***B****) Histogram and kernel density estimation of the ARIMA residual errors.* ***C****) Theoretical vs sample residual error quantile plot.* ***D****) Autocorrelation plot of time series residuals and lagged time series residuals. 95% confidence interval shown in blue.*

**ARIMA model fit evaluation.** Figure 6A depicts the residual errors of the ARIMA model predictions for all predicted days. The residual errors exhibit uniform variance and fluctuate generally between -2 and 2 around a mean of zero. This indicates that the model predictions are reliable.

Figure 6B shows the density plot describing the distribution of: the residual errors shown in blue bars and the kernel density estimation (KDE) shown in orange. In green, a normal distribution curve with mean centered at 0 is shown. Both the shape of the residual errors and KDE indicate a mean of 0 and they closely resemble the normal distribution which indicates robust model performance.

Figure 6C exhibits all the data points, reflecting residual errors, falling in line with the diagonal, shown in red. No significant deviations from the diagonal suggests that the distribution does not reveal any skewness.

Finally, Figure 6D illustrates the correlogram, or the autocorrelation function plot, which shows that the residual errors are within the 95% confidence interval and therefore are not autocorrelated. Lack of autocorrelation of residual errors implies that there is no pattern in the residual errors which is not explained by the model, indicating that no more instances should be added to the model.

## 4 Discussion

Our trained ARIMA and ML models outperformed the basic, single-day and single-feature predicting baseline predictions for subsequent day mood prediction, suggesting that the models successfully captured the predicting capabilities of the features. Both models exhibit a small range in their error metrics (4% for SVR and 3% for ARIMA) indicating no significant interference from outlying predictions. ARIMA offers a greater improvement in prediction against the baseline compared to SVR. ARIMA had smaller errors compared to SVR indicating potentially spurious superior performance. This could be explained by the fact that ARIMA is specifically built to work with time series, while excessive preprocessing is needed with classic ML models. However, the two models are not directly comparable, as different preprocessing was needed prior to model building and training. The discrepancy in error is likely a consequence of differential data preprocessing rather than just actual model performance.

A major drawback in our experiment was the availability of limited data, causing downstream limitations in model generalisation and reliability. Preprocessing for the ARIMA model resulted in a large decrease in dataset size to 98 instances which led to two main drawbacks: using a small training set (from 1-68) for each prediction and not having enough data to forecast personalised mood. Another downside of the univariate ARIMA approach is the sole consideration of a single variable, mood, to make predictions. Although this feature acts as the best predictor, some predictive power from additional features is consequently lost. Conversely, the SVR approach integrated the predictive power of a subset of features (textitmean_mood, sum_screemean_activity, mean_circumplex.valence, sum_appCt.office, sum_appCat.communication, sum_appCat.entertainment, sum_appCat.builtin, sum_appCat.other, sum_appCat.travel, sum_sms, mean_circumplex.arousal, sum_appCat.finance, sum_appCat.social). A subset of them is also validated as predictors by the literature due to their effect on mood, assuring the effectiveness of the feature selection process [16, 23, 9, 19]. The SVR approach also reduced the dataset size, albeit not as dramatically as ARIMA, to 872 instances. However, this was enough to prohibit personalised subject mood prediction. Personalised mood prediction is desirable. However, significant amounts of personal data are needed. Personalised mood prediction was inferior to non-personalised mood with limited data per subject, although personalised predictions exhibited a positive trend speculated to outperform non-personalised if more data were available [3].

The inability to predict extreme mood values might be due to the average computed during windowing, which reduces variance. A solution to this could be to decrease the window size. Mood is itself a convoluted attribute, as it is the amalgamation of several variables and it is influenced by many biological, environmental and lifestyle factors, which are challenging to quantify and include into a predictive model [7, 13].

Consequently, reductive measures are sampled which are distantly related to mood. These crudely, vaguely or erroneously describe the target mood prediction, especially in their raw form. The dataset utilised the readings of an accelerometer as a proxy for physical activity, which resulted in questionable and potentially spurious relationships between the variables measured, such as higher correlation in time spent on sum_appCat.communication and sum_screen, compared to mean_circumplex.valence. These findings are contradictory to the literature investigating the effect of exercise in influencing emotion in a putatively comparable population, and highlight a limitation of the data collection [8]. Furthermore, self-reporting is generally unreliable, as it is retrospective and subjects rely on heuristics when self-reporting [15]. The subject population self-reported relatively positive mood frequently throughout the experiment duration (mean=7), which biases our models towards higher mood value prediction and very short bouts of low mood, making them less likely to generalise well in a real-life scenario. Furthermore, this dampens the range of variance, thereby hindering feature selection algorithms from selecting effective predictors [3, 22].

The biggest limitation to our model performance was the dataset size. Longer data sampling periods per subject would enable more accurate predictions and permit personalised future mood predictions [3]. Furthermore, there is a difference between the sexes regarding problematic smartphone use, and therefore it would be pertinent to further investigate this disparity [25]. Improvements in the model building could have been implemented, such as using a different data aggregation method for the ARIMA dataset or implementing a CV approach to determine optimal size for the sliding window in the windowing process. However, due to limited time these were not feasible. Collection of more mood-relevant attributes (e.g. sleep duration, daily distance walked) or further processing of features in the dataset to more meaningful attributes could improve model performance [3]. Utilising elaborate models and multiple mood-relevant features enabled by multivariate ARIMA, for instance, or using of a combination of models or neural networks could vastly improve predictive performance.

Our results support our hypothesis that smartphone readings and behaviours are able to inform about and predict future moods. Although not perfectly, models can predict the general improvement or worsening of mood better than our baseline. These findings back the notion that wearable and smartphone data could immensely inform and contribute to human well-being, when positive-mood-promoting behaviours are identified. The use of more elaborate and detailed models, such as neural networks, could greatly improve model predictive performance and plausibly be used in mental health therapeutic contexts.

## References


[1] Aljohara A Alhassan et al. "The relationship between addiction to smartphone usage and depression among adults: a cross sectional study". In: *BMC psychiatry* 18.1 (2018), pp. 1–8.

[2] Laith Alzubaidi et al. "Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions". In: *Journal of big Data* 8.1 (2021), pp. 1–74.

[3] Joost Asselbergs et al. "Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: an explorative study". In: *Journal of medical Internet research* 18.3 (2016), e5505.

[4] Mariette Awad and Rahul Khanna. "Support vector regression". In: *Efficient learning machines*. Springer, 2015, pp. 67–80.

[5] Olusola Ayandele, Olugbenga Popoola, and Tolulope Oladiji. "Addictive use of smartphone, depression and anxiety among female undergraduates in Nigeria: a cross-sectional study". In: *Journal of Health Research* ahead-of-print (Apr. 2020). DOI: `10.1108/JHR-10-2019-0225`.

[6] L Ceci. *Average time spent daily on a smartphone in the United States 2021*. 2022. URL: `https://www.statista.com/statistics/1224510/time-spent-per-day-on-smartphone-us/#:~:text=According%5C%20to%5C%20a%5C%20survey%5C%20conducted,average%5C%20on%5C%20their%5C%20phone%5C%20daily` (visited on 04/18/2022).

[7] Corrado Garbazza and Francesco Benedetti. "Genetic factors affecting seasonality, mood, and the circadian clock". In: *Frontiers in endocrinology* 9 (2018), p. 481.

[8] Grace E Giles et al. "Endurance exercise enhances emotional valence and emotion regulation". In: *Frontiers in human neuroscience* (2018), p. 398.

[9] Google. *Uncovering Digital Wellbeing in EMEA*. 2019. URL: `https://www.thinkwithgoogle.com/feature/digital-wellbeing-statistics/` (visited on 04/21/2022).

[10] Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. OTexts, 2018.

[11] Maidul Islam. "Link between Excessive Smartphone Use and Sleeping Disorders and Depression among South Korean University Students". In: *Healthcare*. Vol. 9. 9. Multidisciplinary Digital Publishing Institute. 2021, p. 1213.

[12] John D Kelleher, Brian Mac Namee, and Aoife D'arcy. *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. MIT press, 2020.

[13] Julian Kiverstein, Mark Miller, and Erik Rietveld. "How mood tunes prediction: a neurophenomenological account of mood and its disturbance in major depression". In: *Neuroscience of Consciousness* 2020.1 (2020), niaa003.

[14] Federica Laricchia. *End-user spending on wearable devices worldwide from 2019 to 2022(in billion U.S. dollars)*. 2022. URL: `https://www.statista.`

com/statistics/1065284/wearable-devices-worldwide-spending/ (visited on 04/18/2022).

[15] Richard E Lucas. "Reevaluating the strengths and weaknesses of self-report measures of subjective well-being". In: *Handbook of well-being. Salt Lake City, UT: DEF Publishers* (2018).

[16] KC Madhav, Shardulendra Prasad Sherchand, and Samendra Sherchan. "Association between screen time and depression among US adults". In: *Preventive medicine reports* 8 (2017), pp. 67–71.

[17] Jocelyne Matar Boumosleh and Doris Jaalouk. "Depression, anxiety, and smartphone addiction in university students-A cross sectional study". In: *PloS one* 12.8 (2017), e0182239.

[18] Mani Mehraei and N Akcay. "Mood and emotional states prediction by time series methods". In: *International Psychological Applications Conference and Trends. Lisbon: Research Gate.* 2016.

[19] Adam Mikus et al. "Predicting short term mood developments among depressed patients using adherence and ecological momentary assessment data". In: *Internet interventions* 12 (2018), pp. 105–110.

[20] Christian Reiser. "Predicting and Visualizing Daily Mood of People Using Tracking Data of Consumer Devices and Services". In: *arXiv preprint arXiv:2202.03721* (2022).

[21] Akane Sano et al. "Prediction of happy-sad mood from daily behaviors and previous sleep history". In: *2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE. 2015, pp. 6796–6799.

[22] Jerome Sarris et al. "Multiple lifestyle factors and depressed mood: a cross-sectional and longitudinal analysis of the UK Biobank (N= 84,860)". In: *BMC medicine* 18.1 (2020), pp. 1–10.

[23] Vera Shuman, David Sander, and Klaus R Scherer. "Levels of valence". In: *Frontiers in psychology* 4 (2013), p. 261.

[24] Laura Silver. *Smartphone Ownership Is Growing Rapidly Around the World, but Not Always Equally.* 2019. URL: https://www.pewresearch.org/global/2019/02/05/smartphone-ownership-is-growing-rapidly-around-the-world-but-not-always-equally/ (visited on 04/18/2022).

[25] Sei Yon Sohn et al. "Prevalence of problematic smartphone usage and associated mental health outcomes amongst children and young people: a systematic review, meta-analysis and GRADE of the evidence". In: *BMC psychiatry* 19.1 (2019), pp. 1–10.

[26] Yehuda Wacks and Aviv M Weinstein. "Excessive smartphone use is associated with health problems in adolescents and young adults". In: *Frontiers in psychiatry* 12 (2021), p. 762.