Bayesian Networks for Mood Prediction Using Unobtrusive Ecological Momentary Assessments

Rebolledo, M.; Eiben, A.E.; Bartz-Beielstein, T.

published in
Applications of Evolutionary Computation
2021

DOI (link to publisher)
10.1007/978-3-030-72699-7_24

document version
Publisher's PDF, also known as Version of record

Bayesian Networks for Mood Prediction Using Unobtrusive Ecological Momentary Assessments

Margarita Rebolledo[1(✉)], A. E. Eiben[2], and Thomas Bartz-Beielstein[1]

[1] Institute for Data Science, Engineering and Analytics, TH Köln, Cologne, Germany
{margarita.rebolledo,thomas.bartz-beielstein}@th-koeln.de
[2] Department of Computer Sciences, Vrije Universiteit Amsterdam, Amsterdam, Netherlands
a.e.eiben@vu.nl

Abstract. Depression affects an estimated 300 million people around the globe. Early detection of depression and associated mental health problems constitutes one of the best prevention methods when trying to reduce the disease's incidence. Information collected by tracking smartphone use behaviour and using ecological momentary assessments (EMA) can be used together with machine learning techniques to identify patterns indicative of depression and predict its appearance, contributing in this way to its early detection. However many of these techniques fail to identify the importance and relationships between the factors used to reach their prediction outcome. In this paper we propose the use of Bayesian networks (BN) as a tool to analyse and model data collected using EMA and smartphone measured behaviours. We compare the performance of BN against results obtained using support vector regression and random forest. The comparison is done in terms of efficacy, efficiency, and insight. Results show that no significant difference in efficacy was found between the models. However, BN presented clear advantages in terms of efficiency and insight given its probability factorization, graphical representation and ability to infer under uncertainty.

Keywords: Bayesian networks · Modelling · Ecological momentary assessments · Interpretability

1 Introduction

Depression and other associated mental health disorders can have disturbing effects on every life aspect of people suffering from it. According to the world health organization an estimated 300 million people were suffering from depression in 2015. It is also the largest contributor to global disability. Prevention and early recognition can reduce the incidence of depression and mitigate the negative impacts associated with this disorder.

Several studies have linked the influence of smartphones in the development and intensity of depression and anxiety symptoms [8]. However, they can also

P. A. Castillo and J. L. Jiménez Laredo (Eds.): EvoApplications 2021, LNCS 12694, pp. 373–387, 2021.
https://doi.org/10.1007/978-3-030-72699-7_24

work as a tool for early recognition of depressive indicators. One possible approach is through the use of ecological momentary assessments (EMA) to help evaluate the current mental state of the user. EMA methods allow the collection of psychological phenomena in real-time and within a subject's natural environment to avoid retrospective biases. One common data collection method is self-assessment questions through the user's mobile phone. Another less obstructive approach is making use of software monitoring or embedded smartphone sensors to follow the user's activities and behaviors. The identification of certain patterns could lead to the early detection of depressive symptoms.

Bayesian networks (BN) is a probabilistic graphical model that represents the conditional dependencies between a set of variables. They are especially useful when dealing with restricted amounts of data and in aiding human decision making. According to a recent review of BN use in healthcare [14], its implementation in psychological and psychiatric disorders make up the fourth-largest group of applications. However, this number still lags when considering other machine learning approaches and more work still needs to be done to bring more attention to this modelling method. Some works focusing on the implementation of BN in the field of depression include [15], where possible causal relationships between obsessive-compulsive disorder and depression were disclosed. In [6] an ontology model together with BN is used to infer the probability of becoming depressed. Other works have also used smartphone sensing to monitor and study depression and similar mental health disorders [1,9,21]. Different machine learning methods have been implemented for mood prediction. In [10] an autoencoder is used to fill missing data and provide a robust mood prediction. Deep learning is also implemented in [20] to classify mood readings deviating from common relaxed characteristics. In [7] random forest is used for mood prediction using digital log data. L2-regression and support vector machines are used to predict medical regression using PHQ-9 data in [9].

In a background study [4] different machine learning techniques were tested on EMA to predict the mood of users. Their focus was on predicting the current mood of individual participants based on their smartphone measurements. They concluded that machine learning techniques support vector regression (SVR) and random forest (RF), which incorporate all data about a participant, performed better. However, they stated that more work should focus on finding the most relevant attributes influencing the short-term mood of the participants.

The aim of this work is to expand on the findings in [4] and implement BN as a quick to perform and easy to interpret model to predict the current mood of the user. Given the intrinsic properties of BN new information about the attributes that most influence the mood and their conditional dependencies can be gained. This information can be presented to users for them to have a better understanding of their patterns, be used by medical professionals to provide better insight into personalized and general patterns and behaviors that can influence mood, or, be used as a base to design more comprehensive experiments that allow discovering possible causal relationships between variables. In short, the present study seeks to answer the following questions:

RQ-1 How would BN compare to SVR and RF in terms of efficiency, efficacy, and insight when modelling each participant individually?
RQ-2 How would BN compare to SVR and RF in terms of efficiency, efficacy, and insight when generating a general model for all participants?

The paper is structured in the subsequent manner. Section 2 explains the data available for the model building. In Sect. 3 the different methods used to model the data are presented. Experiments are conducted on Sect. 4 and their results are discussed on Sect. 5. Final thoughts and future work are given in Sect. 6.

2 Data

The data set presented here was first described in [2] and used in [4]. The data originates from a pilot study where 27 university students self-monitored their mood for a time frame of 6 weeks. During this time each participant would be prompted by a cellphone application to input their current mood on a scale from 1 to 10, 10 being the best, five times a day. Additional to logging the mood self-assessment information, the mobile application also recorded other different EMA in the background. Information about the duration and frequency of calls, SMS, screen on/off events, applications used, number of images taken with the camera, and activity were recorded. A total of 55 parameters make up the initial data set. An overview of all variables is presented in Table 1.

Table 1. Mood prediction data set. The attributes names correspond to one or more attributes collecting similar information. The number of variables in each group is given in n_Variables.

Attribute name	Explanation	n_Variables	Range
Id	Participant's identification code	1	[1–27]
mood (Target)	Mood scored by the user	1	[1–10]
callc1c - callc5c	Number of calls to top 5 contacts	5	[0–1]
callc1d - callc5d	Duration of calls to top 5 contacts	5	[0–1]
smsc1c - sms5c	Number of SMS to top 5 contacts	5	[0–1]
accelerometer.high	Percentage of high activity time	1	[0–1]
screen.duration	Standardized total screen-on events	1	[–3,3]
screen.n	Standardized frequency screen-on event	1	[–3,3]
app.a1c - app.a5c	Top 5 Apps usage frequency	5	[0–1]
app-a1d - app.a5d	Top 5 Apps usage duration	5	[0–1]
appCat.n	Apps use frequency	11	[0–1]
appCat.sum	Apps usage duration	11	[0–1]
image.n	Number of images taken	1	[0–1]
mood.l1	Standardized mood of yesterday	1	[–3–3]
mood.l2	Standardized mood of day before yesterday	1	[–3–3]

The target variable *mood* is a daily average. Only days with at least one rating of mood are included in the data set. As a result, participants have a different number of effective samples going from 26 up to 40. Variables related to calls, SMS, images taken, and Apps duration and frequency of use were summarized daily and normalized within participants. The variable accelerometer.high represents the percentage of time during the day in which the participant's activity surpassed a threshold of 10 m/s^2. Lastly, variables mood.l1, mood.l2, screen.duration and screen.n are transformed to the standard normal distribution. That is, 99.97% of their values fall between –3 and 3. For more in-depth information about the different variables please refer to [2].

As an additional preprocessing step, variables with near zero variance were removed from the data set. Variables for which the ratio between the frequency of its most common value to its second most common value was less than 15 were removed.

In the end, the data set used to build the models presented in the following sections contains a total of 24 predictors, excluding target and id variables, and 595 observations across all participants are available.

3 Methods

3.1 Bayesian Networks

BN [12,16] are probabilistic models based on directed acyclic graphs (DAG). A DAG structure is given by $\mathcal{G} = (\mathbf{V}, A)$ where $\mathbf{V}$ is the node and A is the arc set. Nodes represent the random variables of interest $\mathbf{V} = \{X_1, X_2, ..., X_n\}$, and arcs represent informational or causal dependencies among the variables. They are quantified as conditional probabilities for each node given its parent nodes. The DAG defines a factorization of the global probability distribution of $\mathbf{V}$ into local probability distributions, one for each node. The local distribution for a random variable $P(X_i)$ is given as the conditional distribution of X_i and its parents Π_{X_i} as $P(X_i|\Pi_{X_i})$. In other words, the local distribution of a node is independent of other nodes given its parents. Following the same method, the global probability distribution of the network defined by $\mathcal{G}$ is

$$P(X_1, ..., X_n) = \prod_{i}^{n} P(X_i|\Pi_{X_i}) \tag{1}$$

where n is the number of nodes. Any probability of interest can be computed from this joint probability.

In order to learn the BN model from the data two steps are required: structure learning and parameter learning. Structure learning entails finding the DAG encoding the dependence structure of the nodes given the data, $P(\mathcal{G}|D)$, where D is the data. This can be guided by expert knowledge if available. In parameter learning, the parameters Θ, that define the local distributions for each node are

estimated. As already mentioned, these distributions are independent and as such the Θ can be computed as

$$P(\Theta|\mathcal{G}, D) = \prod_{i=1}^{n} P(\Theta_i|\Pi_{X_i}, D) \tag{2}$$

Three main approaches for structure learning are possible: constraint-based, score-based, and hybrid [19].

Constraint-based approaches implement conditional independence tests, such as Fischer's Z test, to discover the dependence structure of the data. In score-based approaches, general-purpose optimization techniques like hill-climbing or Tabu search are used to generate candidate DAGs. Each candidate receives a score reflecting its goodness-of-fit which the optimization algorithm tries to maximize. Finally, hybrid algorithms combine both approaches sequentially by first selecting a network skeleton using constraint-based algorithms and then maximizing its score.

Finally, assumptions need to be done about the distribution followed by $\mathbf{X}$. In general, if the data set contains continuous parameters it is assumed that $\mathbf{X}_i$ follows a multivariate normal distribution and that the relationship with its parents is linear. These networks are known as Gaussian BNs and are the type used in this work.

For the implementation of BN we used the *bnlearn* R package [18]. The network structure is learned via the score-based hill-climbing search by maximizing the Bayesian information criterion (BIC).

3.2 Support Vector Regression

Support vector regression (SVR) [3] is a generalization of the well known support vector machines (SVM) for classification. In the classification case SVM finds the optimal hyperplane separating different categories. In the general case SVR introduces an ε-insensitive region, called ε-tube. The regression is then formulated as an optimization problem where a convex ε-insensitive loss function needs to be minimized to find the flattest tube that contains most of the training data samples. The optimization problem is solved using numerical optimization algorithms. For non-linear functions, the data can be mapped into a higher dimensional space using a kernel function.

To follow the implementation in [4], we implemented SVR on the *kernlab* R package [11]. The SVR was defined as an epsilon regression with epsilon and the cost of constraint violation set to 0.5. The radial kernel was used with hyperparameters values calculated internally by the included heuristic sigest.

3.3 Random Forest

Random Forest (RF) [5] is a well-known type of non-linear multiple regression. It is an ensemble method that groups several weak learners (decision trees) and combines them to generate a strong learner (a forest). For each decision tree, an

input is split into smaller subsets until it no longer brings an improvement to the model response. Another quality of RF is their computation of relative variable importance by measuring the mean decrease in mean square error (MSE) that each parameter generates.

Following the implementation in [4], the number of trees for our models is set to 500 and the number of variables sampled at the splits is set internally by the model. The random forest model is implemented using the R package randomForest [13].

4 Experiments

Two types of experiments are carried out: in the first experimental case each user's data is modeled separately and results will apply only to the specific user. For the second experiment type, the data of all users is combined to build a general model. This will allow examining whether for this data set it is possible to use knowledge of other individuals to improve the prediction of *mood*. The results of the different models on both experiments are compared in terms of efficiency, efficacy, and insight.

The *efficacy* of the models is evaluated in terms of MSE and stability. The average MSE after 10 cross-validations runs is taken as the final MSE of the model. To determine whether a significant difference between the different model's MSE is present the Kruskal-Wallis rank-sum test is used. If the test is positive, a post-hoc test according to Conover [17], for pairwise multiple comparisons, checks for differences in each algorithm pair. The comparisons are further used to rank the algorithms from 1 (best) to 3. The second efficacy measure corresponds to the stability of the model. To control no large changes occur in the prediction given slight changes in the input data, we define stability as the difference between the minimum and maximum result in the model across the 10 cross-validation runs.

The *efficiency* of the models is given as the wall clock time in seconds required to build the models on one machine.

Lastly, the *insight* is understood as the model interpretability. We define interpretability as the informativeness and intelligibility of the model. A model is informative when it allows to explore the data and provide assistance to a human decision maker. Intelligibility is defined as how intuitive the understanding of the model is.

4.1 Individual Models

The first round of experiments is focused on predicting the current *mood* using only the data available for each of the participants. On the one hand, this ensures a completely personalized analysis and acknowledges the differences between each user, but on the other hand, this suffers from a lack of sufficient data samples that some participants presented.

The model validation was done through 10 fold cross-validation. Each participant's data was partitioned into training and testing sets with ratio 85/15. The wall clock time in seconds required by each algorithm to finish all 10 cross-validation runs across all 27 participants is taken. Table 2 presents the results obtained for MSE, time, stability, and rank for the case of individual models. It is clear that all models present the same level of performance, indicated by them all obtaining the same best rank by the post-hoc test. Their only significant difference is the time required to build the model.

Table 2. Individual model results. MSE corresponds to the average across the 10 cross-validation runs and 27 participants. The time in seconds is measured as the total time taken for the 10 repetitions of the 27 models. The rank is assigned after following a post-hoc test according to Conover [17] and rated from 1 (best) to 3.

	SVM	BN	RF
MSE	0.41	0.47	0.39
Stability	1.00	0.87	0.81
Time	30	9	13
Rank	1	1	1

To give a qualitative measure of interpretability the models should be examined individually. As an illustration example, we will use the model for the participant with *id* "AS14.0".

Neither SVR or RF offer an understandable visual depiction of their model structure. The higher dimensionality kernel representation is prohibiting for the case of SVR and the high number of trees in RF is restrictive for the complete representation of the model. In the case of BN the visual representation of the model is more straightforward thanks to its DAG structure.

Figure 1 shows the network learned from user "AS14.01". The arrows' direction indicate the conditional dependencies of the data. For this specific model, *mood* is conditionally dependent of it parents *app.a5c* and *screen.n* such that $P(mood) = P(mood|app.a5c, screen.n)$. The gray shaded nodes depict the Markov blanket of *mood*, or in other words, the subset of variables that have all the information required to compute the probability of the variable of interest according to the joint probability factorization explained in the section above. Also of interest is the arc strength which gives a measure of confidence for each arc. Strength is measured as the decrease in the network score that would be caused by the arcs removals. If it is of interest the network could be reduced using this information. The arc's thickness in Fig. 1 represents the arc's strength, in this case, thicker arcs have a higher confidence level.

A similar functionality to the arcs' strength is given by RF measure of importance. Here the decrease of accuracy, as measured by the MSE, when a variable is removed from the model is taken as the importance of the variable. The parameters importance according to the random forest model of user "AS14.01" is

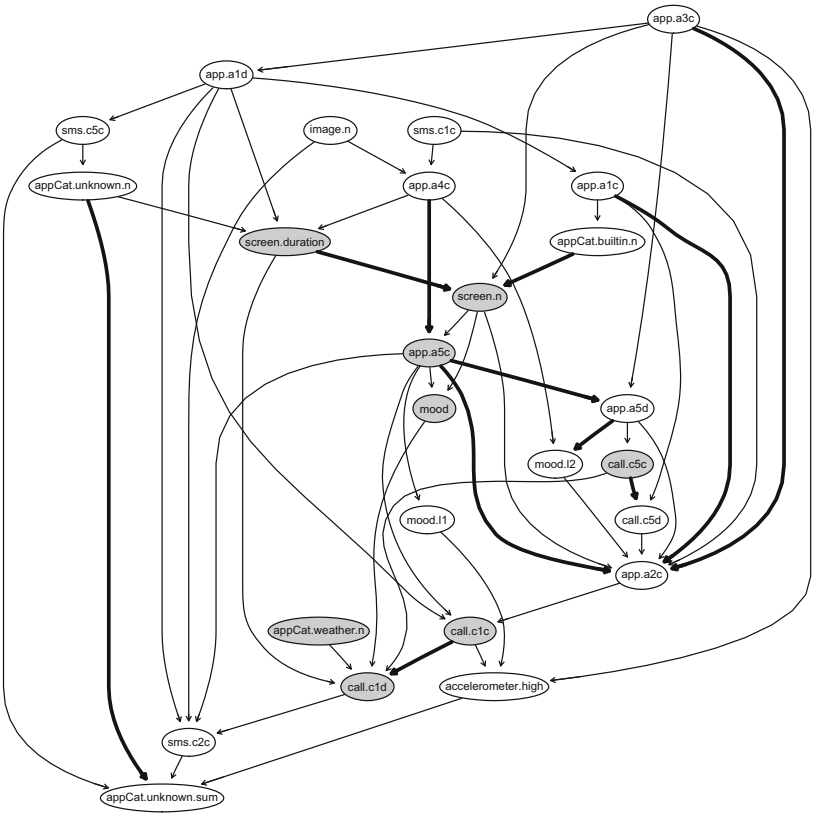

Fig. 1. Learned Bayesian network for user *id* "AS14.01". Arrows represent the conditional probability dependencies of each node. The arcs thickness corresponds to the arc's importance in the network according to their BIC score. The target node *mood* and the nodes conforming its Markov blanket are gray shaded. In this model it can be seen that *screen.n* influences the value of *mood* while *sms.c2c* has no relationship with it.

shown in Fig. 2. It is interesting to note that parameters with higher importance are not all included in the Markov blanket of the BN model.

As a final characteristic of BN that largely impact their interpretability is their ability of inference under uncertainty. It is possible to investigate the effects of new evidence using the information encoded in the BN and compute the maximum posterior density, or simply put, the probability of a given combination of events on a node. If we are interested in finding the probability that for user "AS14.01" a *mood* higher than 6 happens if the *screen.duration* value is higher than 2 (higher than the mean) and the *app.a5c* is lower than 0.03 (lower than the mean) then the probability of *mood* will be $P(mood > 6|screen.duration > 2 \cup app.a5c < 0.03) = 0.96$. This probability decreases if we consider a lower value of *screen.duration* $P(mood > 6|screen.duration < 2 \cup app.a5c < 0.03) = 0.9$.

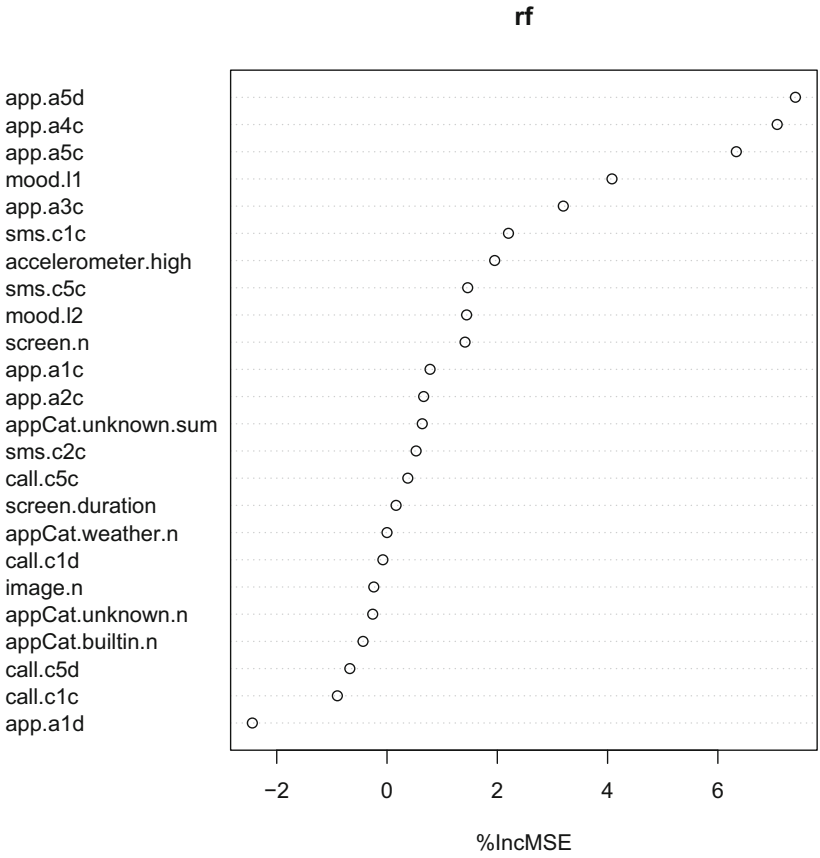

Fig. 2. Random forest importance measure for user *id* "AS14.01". The importance is measured as the decrease of accuracy (MSE) in out of bag samples when the variable is excluded from the model.

This ability of 'what if' queries greatly increase the amount of information that can be obtained from the model

To illustrate the variability between the networks obtained from different participants, we present the learned Bayesian network for user "AS14.31" in Fig. 3. Markov blanket and arc strength are visualized as previously explained. We can see that variables part of the Markov blanket for user "AS14.01", as is the case with *call.c5c*, are considered here conditional independent. Predicted values for these variables are taken from their assumed prior distributions.

4.2 General Model

In the case that behaviors between participants do not differ considerably, it should be possible to combine all the information the data can give and use it to predict *mood*.

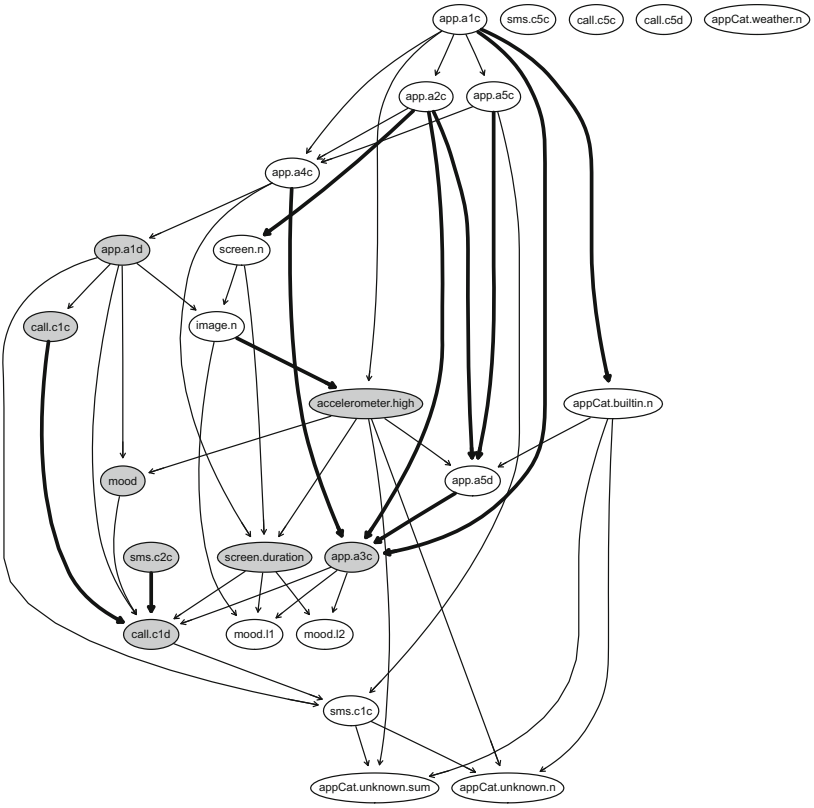

Fig. 3. Learned Bayesian network for user *id* "AS14.31". Arrows represent the conditional probability dependencies of each node. The arcs thickness corresponds to the arc's importance in the network according to their BIC score. The target node *mood* and the nodes conforming its Markov blanket are gray shaded. In this model it can be seen that *app.a1d* influences the value of *mood* while *call.c5c* has no relationship with it.

The same procedure as for the individual models was followed to build the general model. The data is partitioned into test and training sets, and the model was validated using 10 cross-validation runs.

As an initial approach, the models were built omitting any personal information given here by the *id* parameter. Results indicate that despite counting with more data samples no reasonably conditional dependence was established for the *mood* parameter. A look into the performance of all three models indicates that no improvement was seen in the MSE.

The model was build again including the *id* parameter. In this case, the model is able to identify conditional dependencies for *mood*. Results for this general model can be seen in Table 3. Also in this case the models do not present any improvement in performance and no significant difference between them. However, the time saving achieved by BN is notable.

Table 3. Results for the general model. MSE corresponds to the average across the 10 cross-validation runs. The time in seconds is measured as the total time taken for the 10 repetitions. The rank is assigned after following a post-hoc test according to Conover [17] and rated from 1 (best) to 3.

	SVM	BN	RF
MSE	0.45	0.44	0.41
Stability	0.41	0.34	0.38
Time	22	2	25
Rank	1	1	1

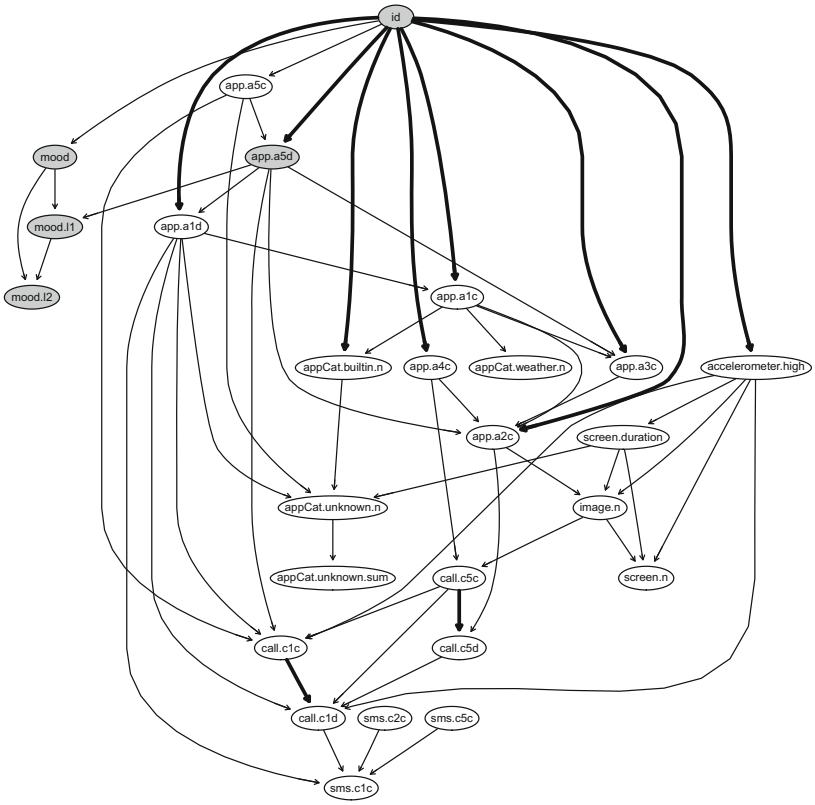

Fig. 4. Bayesian network of the general model. Arrows represent the conditional probability dependencies of each node. The arc's thickness corresponds to the arc's importance in the network according to their BIC score. The target node *mood* and the nodes conforming its Markov blanket are gray shaded. It can be seen that many of the most important dependencies relate to *id*, and *mood* is also conditionally dependent of it. On the other side, there is total independence from *sms.c5c*.

Similar to the individual models' case, the structure of the general model can be observed in Fig. 4. As previously explained, the thicker arcs represent the ten connections with the highest confidence level. As it can be seen in the figure most of them are related to the *id*.

A look into RF measure of importance also shows that *id* is by far the most important parameter influencing the MSE. Taking into account that the model efficacy did not improve with the addition of more data samples and that the identification parameter seems to be of great importance to the understanding of the model, it can be assumed that the prediction of mood using all the participants is not a good idea for this data set.

5 Discussion

In order to test the efficacy, efficiency, and insight of BN against RF and SVR when predicting the current *mood* two types of experiments were carried out.

In the case of individual models, it could be proven that the performance achieved by BN is at the same level of RF and SVR. We also argued that the interpretability of the network, in terms of informativeness and intelligibility, is higher than for the other models. The results also showed that the difference in time required by the algorithms is noticeable.

It is important to consider that students presenting symptoms of clinical depression were excluded from the data collection process generating a bias in the *mood* data. More tests are needed with new data where the presence of depression is given to explore the changes in relationships between parameters.

BN are probabilistic models that can portray relationships of causality between its parameters. However, causality cannot be confirmed and only a conjectural case for causality can be started when referring to the models presented in this work. More data needs to be obtained preferably from controlled experiments that account for confounding or latent variables.

For the general model case, it was observed that even though the number of data samples increased, no improvement in the model accuracy was observed. The lack of increase in performance given the increase in the number of data points led us to conclude that the model was not able to obtain more information about the participants and thus for the given data set it is not appropriate to generate a general model that can combine information about all patients. The addition of new parameters or a wider range of *mood* values could be necessary to achieve a good model.

The similarity in the performance of the tested models make criteria such as interpretability an important concept to define according to the end purpose of the model. In a scenario of medical studies where analysis should be personalized for each patient, it is beneficial to have a model that allows the medical professional to understand the flow of data and its conditional dependencies to plan experiments and better analyze results.

6 Conclusion

The use of unobtrusively collected EMA via smartphone as a way to predict the current mood of the user through BN is studied in this work. To answer our research questions we compared the efficacy, efficiency, and insight of BN against the already tested models SVR and RF. The efficacy is measured in terms of MSE and stability. The wall-clock time measures the efficiency and the level of insight is measured in terms of interpretability. Two types of experiments were carried out on the data in order to answer our research questions.

RQ-1 How would BN compare to SVR and RF in terms of efficiency, efficacy, and insight when modelling each participant individually? In our first experiment, one model per participant was build and the performance of the models was evaluated using cross-validation. The results showed that no significant difference in performance was found between the models. In terms of speed and interpretability BN presented clear advantages. Interpretability is defined as the informativeness and intelligibility of a model. In Fig. 1, it was easy to identify which of the parameters had an influence on the outcome of *mood* based on the arc connections present between the nodes and the visual representation of its Markov blanket. At the same time, it was easy to represent the conditional probabilities with the most importance in the model.

RQ-2 How would BN compare to SVR and RF in terms of efficiency, efficacy, and insight when generating a general model for all participants? For the second type of experiment one general model was implemented. Results were consistent in terms of performance between the three models. Also here the gains in interpretability and speed were noticeable for BN. However, it was clear given the lack of performance improvement that the amount and quality of data may be insufficient to build such a general model and neither of the tested methods presented a considerable improvement.

BN as a tool to graphically model the conditional dependencies between variables do not present any loss in performance against SVR or RF. On the other hand, it was shown that BN bring benefits in terms of efficiency and insight given its probability factorization, graphical representation, and ability to infer under uncertainty. These advantages would make its implementation as informative models in mobile devices realizable. A conceivable application would be a weekly presentation of the built model to inform the user of their behavioral trends.

In future work, we would like to make use of another important feature of BN that allows the inclusion of expert knowledge into the model. In this way, relationships between already studied parameters can be included and inference and reasoning capabilities of the model can be improved. More work is also needed using a more comprehensive data set in order to achieve a more robust understanding of the influence certain behaviors have on the mood. It is necessary to evaluate the inclusion of new measurable variables in the data collection, such as sleep duration and patterns. Feature selection and feature engineering to define the parameters to be modeled remain as an open possibility that needs to be explored.

References

1. Aledavood, T., Torous, J., Triana Hoyos, A.M., Naslund, J.A., Onnela, J.P.,Keshavan, M.: Smartphone-based tracking of sleep in depression, anxiety, and psychotic disorders. Current Psychiatry Rep. **21**(7), 49 (2019).https://doi.org/10.1007/s11920-019-1043-y
2. Asselbergs, J., Ruwaard, J., Ejdys, M., Schrader, N., Sijbrandij, M., Riper, H.: Mobile phone-based unobtrusive ecological momentary assessment of day-to-day mood: An explorative study. J. Med. Int. Res. **18**(3), e72–e72 (2016). https://doi.org/10.2196/jmir.5505
3. Awad, M., Khanna, R.: Support Vector Regression, pp. 67–80. Apress, Berkeley, CA (2015). https://doi.org/10.1007/978-1-4302-5990-9_4
4. van Breda, W., Pastor, J., Hoogendoorn, M., Ruwaard, J., Asselbergs, J., Riper, H.: Exploring and comparing machine learning approaches for predicting mood over time. In: Chen, Y.-W., Tanaka, S., Howlett, R.J., Jain, L.C. (eds.) Innovation in Medicine and Healthcare 2016. SIST, vol. 60, pp. 37–47. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-39687-3_4
5. Breiman, L.: Random forests. Mach. Learn. **45**(1), 5–32 (2001). https://doi.org/10.1023/A:1010933404324
6. Chang, Y.S., Fan, C.T., Lo, W.T., Hung, W.C., Yuan, S.M.: Mobile cloud-based depression diagnosis using an ontology and a Bayesian network. Future Gener. Comput. Syst. **43-44**, 87–98 (2015).https://doi.org/10.1016/j.future.2014.05.004, http://www.sciencedirect.com/science/article/pii/S0167739X1400137X
7. Cho, C.H., Lee, T., Kim, M.G., In, H.P., Kim, L., Lee, H.J.: Mood prediction of patients with mood disorders by machine learning using passive digital phenotypes based on the circadian rhythm: prospective observational cohort study. J. Med. Int. Res. **21**(4), e11029 (2019). https://doi.org/10.2196/11029, http://www.ncbi.nlm.nih.gov/pubmed/30994461
8. Elhai, J.D., Dvorak, R.D., Levine, J.C., Hall, B.J.: Problematic smartphone use: a conceptual overview and systematic review of relations with anxiety and depression psychopathology. J. Affect. Disord. **207**, 251– 259 (2017). https://doi.org/10.1016/j.jad.2016.08.030, http://www.sciencedirect.com/science/article/pii/S0165032716303196
9. Farhan, A.A., et al.: Behavior vs. introspection: refining prediction of clinical depression via smartphone sensing data. In: 2016 IEEE Wireless Health (WH), pp. 1–8 (2016). https://doi.org/10.1109/WH.2016.7764553
10. Jaques, N., Taylor, S., Sano, A., Picard, R.: Multimodal autoencoder: a deep learning approach to filling in missing sensor data and enabling better mood prediction. In: 2017 Seventh International Conference on Affective Computing and Intelligent Interaction (ACII), pp. 202–208 (2017). https://doi.org/10.1109/ACII.2017.8273601
11. Karatzoglou, A., Smola, A., Hornik, K., Zeileis, A.: kernlab - an S4 package for kernel methods in R. J. Stat. Softw. **11**(9), 1–20 (2004). http://www.jstatsoft.org/v11/i09/
12. Korb, K.B., Nicholson, A.E.: Bayesian Artificial Intelligence. CRC Press, Boca Raton (2010)
13. Liaw, A., Wiener, M.: Classification and regression by random forest. R News **2**(3), 18–22 (2002). https://CRAN.R-project.org/doc/Rnews/

14. McLachlan, S., Dube, K., Hitman, G.A., Fenton, N.E., Kyrimi, E.: Bayesian networks in healthcare: distribution by medical condition. Artif. Intell. Med. **107**, 101912 (2020).https://doi.org/10.1016/j.artmed.2020.101912, http://www.sciencedirect.com/science/article/pii/S0933365720300774
15. McNally, R., Mair, P., Mugno, B., Riemann, B.: Co-morbid obsessive-compulsive disorder and depression: a Bayesian network approach. Psychol. Med. **47**(7), 1204 (2017). https://www.mcnallylab.com/wp-content/uploads/McNally-Mair-Mugno-Riemann-2017-Psychological-Medicine.pdf
16. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufmann Publishers Inc., San Francisco (1988)
17. Pohlert, T.: The pairwise multiple comparison of mean ranks package (pmcmr) (2014). http://CRAN.R-project.org/package=PMCMR. Accessed 12 Jan 2016
18. Scutari, M.: Learning Bayesian networks with the bnlearn R package. J. Stat. Softw. **35**(3), 1–22 (2010). https://doi.org/10.18637/jss.v035.i03, https://arxiv.org/pdf/0908.3817.pdf
19. Scutari, M., Graafland, C.E., Gutiérrez, J.M.: Who learns better Bayesian network structures: accuracy and speed of structure learning algorithms. Int. J. Approximate Reasoning **115**, 235–253 (2019). https://doi.org/10.1016/j.ijar.2019.10.003, http://www.sciencedirect.com/science/article/pii/S0888613X19301434
20. Spathis, D., Servia-Rodriguez, S., Farrahi, K., Mascolo, C., Rentfrow, J.: Passive mobile sensing and psychological traits for large scale mood prediction. In: Proceedings of the 13th EAI International Conference on Pervasive Computing Technologies for Healthcare. pp. 272–281. PervasiveHealth 2019, Association for Computing Machinery, New York, NY, USA (2019). https://doi.org/10.1145/3329189.3329213
21. Wahle, F., Kowatsch, T., Fleisch, E., Rufer, M., Weidt, S.: Mobile sensing and support for people with depression: a pilot trial in the wild. JMIR Mhealth Uhealth **4**(3), e111 (2016). https://doi.org/10.2196/mhealth.5960,http://www.ncbi.nlm.nih.gov/pubmed/27655245