Aspects of a learned model to predict the quality of life of university students in Brazil

Pedro Francis Lopes[1] **, Amilton Santos Junior**[2]**, Renata Cruz Soares de Azevedo** [2] **,**
Paulo Dalgalarrondo [2] **, Andre Ricardo Fioravanti**[1]

[1]UNICAMP - University of Campinas, School of Mechanical Engineering

[2]UNICAMP - University of Campinas, School of Medical Sciences

p003265@unicamp.br

Abstract. *Quality of life is an essential metric for evaluating the well-being of students. This work investigates the viability of a model to predict a WHOQoL-Bref answer based on other answers and the overall domain and average scores. For that, we use data from an extensive pooling done with undergraduate students in Brazil (UNICAMP), gathered between 2017 and 2018. We also discuss model types and hyperparameter effects on model evaluation metrics. Finally, we conclude that it is possible to create a model to predict the esteem question - which is the most correlated with the average domain score with the data sample available.*

1. Introduction

Quality of Life (QoL), as defined by the World Health Organization (WHO), is the perception of one's position in life in the context of the value systems and culture in which they live and their expectations, goals, standards and concerns. It is a subjective evaluation connected to the cultural, social and environmental context [17].

One way to measure the quality of life is to use the WHOQOL-Bref instrument [5]. It has been employed in several applications, including to assess medical students [11]. Undergraduate students are a particular group of interest in terms of QoL for the pressures and stresses they suffer as their lives change and they enter adulthood [3].

The WHOQOL-Bref instrument was chosen because it was embedded in a more extensive questionnaire. Utilizing the complete WHOQOL-100 with more questions would reduce the remaining interview time for other subjects. The WHOQOL manual [5] provides the proper equivalence considerations.

The particular interest in undergraduate university students' psychological and psychiatric features led to a research project conducted at the University in Campinas (UNICAMP) between 2017 and 2018. The project consisted of constructing a questionnaire and pooling the undergraduate students in their various stages at the university. This questionnaire, among other questions, contained the Brazilian version of WHOQOL-Bref for QoL assessment. The instrument was applied during classes on paper and then transcribed into an electronic database. Care was taken to keep the data clean and sanitized. This research yielded 6906 individual feature groups representing each respondent student.

With this database in hand, several opportunities arise for its possible analysis. Besides the traditional and accepted statistical analysis, a trend in using Machine Learning (ML) to analyse student data and medical problems have been gaining traction. For example, [9] presented an interesting paper about the risk of school dropout based on student data. Also, [15] proposed an engaging insight into student performance. On the medical side, the works [16, 6] also present exciting possibilities. Perhaps the most inspiring on the interface of students and psychiatric issues (in particular, Suicide Thoughts and behaviours - STB) is the work presented by Macalli et al. 1 in 2021 [10] where they attempt to predict STB in a database that bears some resemblance to that currently at hand.

There are several ML model types available, which causes a constant struggle to decide the most adequate for each problem. This discussion is still active, both in the data scientist guild and among those intent on automatizing the whole ML model task. The following non-exhaustive list illustrates the variability of methods applied to tackle similar problems:

- Pinto et al. [13] use a Random Forest Regression (RFR) model to predict Health-Related Quality of Life (HRQoL) in melanoma patients.
- Amenomori et al. [1] use the Gradient Tree Boosting (GBT) model to identify QoL impacts on Amyotrophic Lateral Sclerosis (ALS) patient caregivers.
- Kaur et al. [8] use several model types (Decision Tree (DT); Neural Network(NN); Random Forest (RF); Support Vector Regressor (SVR)) to create an ensemble model to predict the Better Life Index (BLI) that is a predictor for QoL.
- Batata et al. [2] use DT, RF and Boosting Classifier (BC) to predict caregivers' burnout.
- Macalli et al. [10] uses RF to predict STB in students.

There is still no consensus on what model type is most adequate for this type of application problem. The choice of type heavily depends on data characteristics and how the features are structured. Therefore, the models listed in the references will be used in the preliminary model training. Thus, this work intends to combine the QoL information in the previously mentioned database with ML models to verify the feasibility of feature prediction. We will explore different model types and hyperparameter searches on selected models, assessing their sensibility to the performance of the model.

2. Materials

The database used in this study was obtained with the paper application of a questionnaire on 6906 undergraduate students of UNICAMP between 2017 and 2018, as mentioned earlier [14, 7]. UNICAMP Research Ethics Committee (IRB - Institution Review Board) granted ethical approval for this study filed under number 1.903.287.(CAAE 62765316.6.0000.5404) The participation was voluntary, and the participant filled out the response sheet in the presence of the researchers during a standard class. The response sheets were then manually transcribed into a database.

The research instrument has 238 questions. There are categorical, numerical and textual features. On the latter, the respondent could explain his/hers feelings or provide further information. In addition, some of the main questions had sub-items that were each treated as a different feature, increasing the overall feature number.

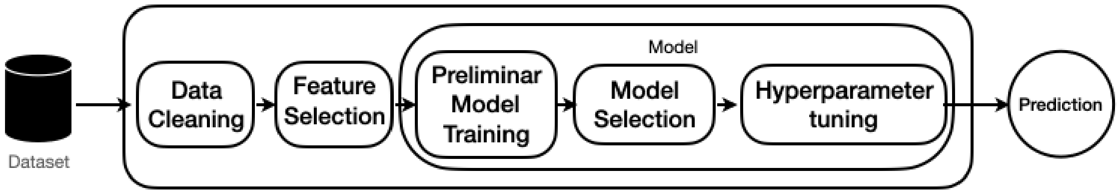

Figure 1. Proposed pipeline topology for the problem.

The questionnaire sections include General information (gender, age, living conditions, parent information, socioeconomic information, student background, perceptions about university and plans for the future); Quality of life (a transcription of the WHOQOL-Bref instrument); Identity; Religion; Physical Activities & Health; Physical Health; Mental Health (that includes an SRQ20); Suicide Thoughts and behaviours (STB); Internet usage; Alcohol & Drug usage; Values and World View; Sexuality and love life; Sexual orientation & other topics.

Pre-processing: Whenever applicable, companion features for the categorical questions were created, indicating intensity or a binary response, e.g. Yes & No, was equated to 1 & 0. The current working database has 6906 instances, 920 features (with 29.2% of missing data) and 120 meta attributes (textual) (with 84% of missing data).

The current work will focus on the WHOQOL-Bref questions within the database. The canonical instrument [5] has 26 questions. The database section regarding QoL, in turn, has 62 features. These features are explained as the categorical question and its accompanying numerical counterpart (52 features). These numerical features range from -2 to 2 (e.g. for satisfaction) or 0 to 4 (e.g. for intensity) whenever appropriate. These values differ from the original 1 to 5 scoring though they bear the same linearity and proportion, just shifting the values. Other numerical/semantic correlations were not the focus of the current study.

In addition, there are instructions on the generation of 4 numerical features - called "Domain Factors", namely "Physical; Psychological; Social Relationships; Environment". These are numerical features by nature. Accompanying categorical features (divided into five equally spaced answers covering 20 points each) were concocted, totalling eight more features. The two additional features are the average over the domain factors (not prescribed on the original material) and its accompanying categorical feature - done in the same manner as the individual domains, thus totalling 62 features. This section of the database has 0.6% of missing elements.

3. Methods

The ML models used are standard, and their implementation can be found on the Sci-kit learn library [12]. For this problem, an ML pipeline is proposed (figure 1), based on a standard pipeline as presented by Zoller et al. [18]. The specifics of this are explained in the following subsections.

3.1. Data Cleaning & Sanity

Data were initially obtained using a paper questionnaire. Humans then transcribed the inputs into a database. Finally, data sanity was verified by checking the questionnaire database by a different group of people.

Features were verified for blanks and spurious data - such as numbers where categories are expected and vice-versa. Once identified, the spurious data was suppressed, and this entry's feature was left blank.

3.2. Feature Selection

For regression models, only the 31 numeric terms were be used; similarly, for the classification models, only the 32 categorical features were used. These features are related to the Quality of Life section of the questionnaire, and additional analysis may be the subject of future work.

The first step was to define the inputs and outputs of the model; in other words, the prediction target and the inputs. A first clear choice was to compare the self-assessment question q58(1) - database number (original instrument equivalent) with the average domain scores.

A more involved solution is to obtain the correlation between the features and the average score. The most correlated feature was chosen as a target, and different models were trained to find the most promising in terms of accuracy, precision and recall.

Since the data size is reasonably manageable with the available computational resources, all features not selected as the target were used on the model as inputs.

3.3. Model Evaluation

Defining true positives (T_P) as a positive outcome correctly classified; true negative (T_N) as a negative outcome correctly classified; false positive (F_P) as a positive outcome incorrectly classified; false negative (F_N) as a negative outcome incorrectly classified. $F1$ is the F-score and can be interpreted as the harmonic mean of precision and recall. The evaluation parameters adopted were:

$$\begin{aligned}
\text{Classification Accuracy} &= \frac{T_P + T_N}{T_P + T_N + F_N + F_P} \\
\text{Precision} &= \frac{T_P}{T_P + F_P} \\
\text{Recall} &= \frac{T_P}{T_P + F_N} \\
\text{F1} &= \frac{T_P}{T_P + \frac{1}{2}(F_P + F_N)} \\
\text{Specificity} &= \frac{T_N}{T_N + F_P}
\end{aligned}$$

4. Results

4.1. Sample Description

WHOQOL Domains The scores were calculated using the given answers and the procedure from WHOQOL-Bref manual [5]

Table 1 shows the mean, median and dispersion of the four domains and the average score. Figure 2 shows the distribution of each part of the WHOQoL-Bef, and 3 shows the average score distribution for each domain.

Table 1. Feature Statistics - WHOQoL Domains

WHOQoL Domain	Mean	Median	Dispersion
Physical	61.4	60.7	0.25
Psychological	56.7	58.3	0.32
Social	60.5	58.3	0.35
Environmental	61.1	62.5	0.26
Overall	59.9	60.5	0.23

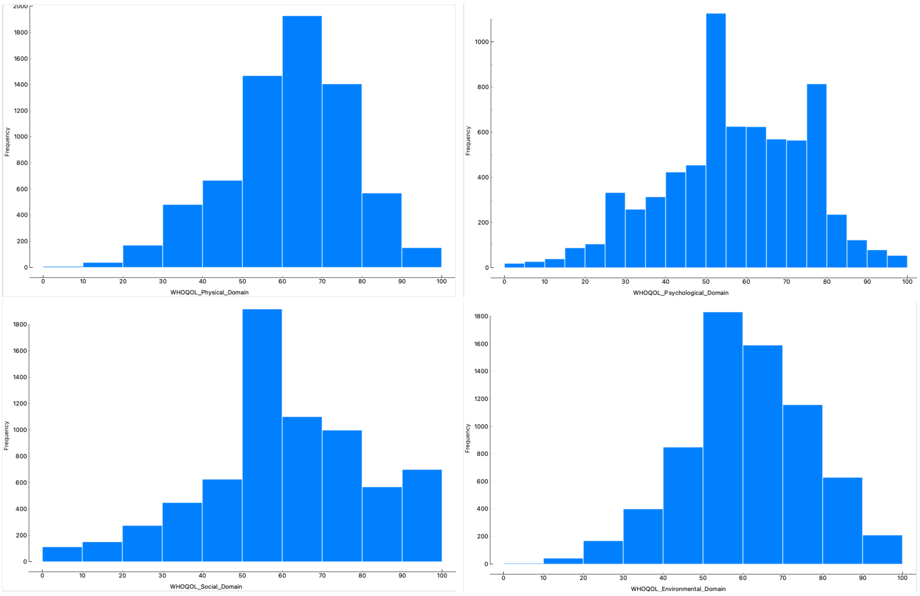

Figure 2. Statistical distribution of each domain. From the top left corner clockwise there are the physical; psychological; social and environmental domains

The distributions for the domain scores may seem normal; however, there is no reason to believe that the underlying process is normal and would generate a Gaussian distribution. Thus, further tests must be done to confirm or deny the normality hypothesis.

The distributions show that the domain values concentrate in the middle of the distribution.

4.2. Self Assessment vs Domain Average

Figure 4; 5 and 6 show the relation between the self-assessment question and the average of the domain scores. The average of the domain score was divided into five boxed of 20 points in size and named ”very low” to ”very high”. This was done to enable a category comparison. The numerical and categorical comparisons can show the same data with different insights.

Though the self-assessment question seems to align with the average of the WHOQoL scores, figure 6 reveals mismatching categories between the self-assessment and the

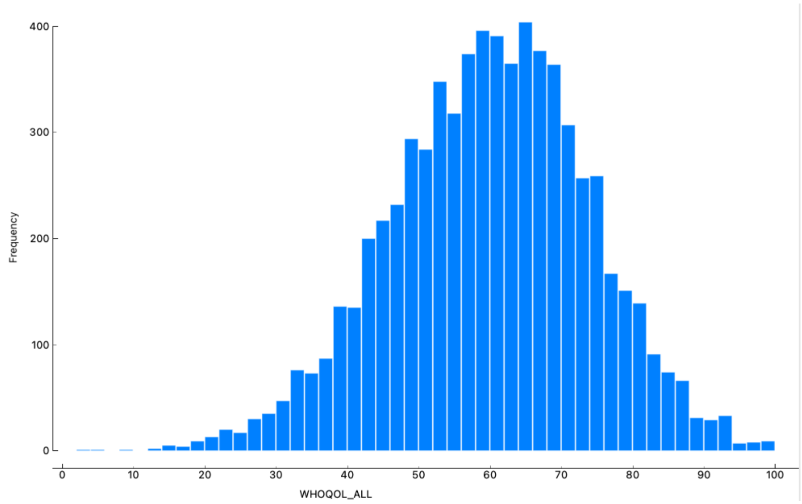

Figure 3. Statistical distribution of the average of the 4 domains.

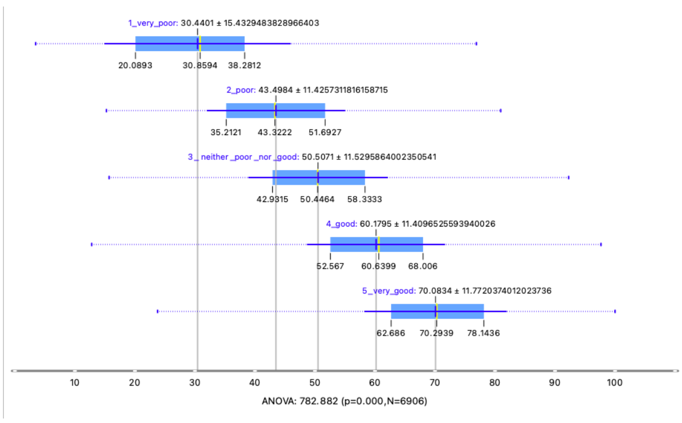

Figure 4. Correlation between the self-assessment question (categorical) and the average of the domain score.

average domain score. This may indicate that the respondents could not correctly self-assess their QoL in the sample.

4.3. Correlations

The correlations presented in this subsection are calculated with the numerical features.

The Pearson correlation between q58(1), the self-assessment question and the average domain score are 0.557 and 0.543 on the Spearman correlation. This partially explains the results from the previous subsection.

Table 2 shows the correlations for the other features.

The highest correlations are the domain themselves, as expected, though not obvious. The question that best correlates with the overall score, apart from the domains, is the esteem question (q76(19)), which seems a good option for the first machine learning model.

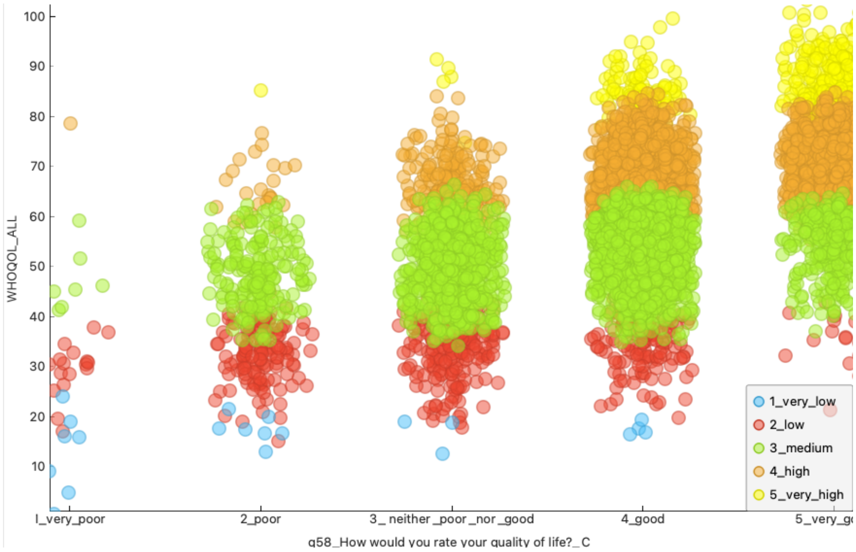

Figure 5. Correlation between the self-assessment question (numerical) and the average of the domain score.

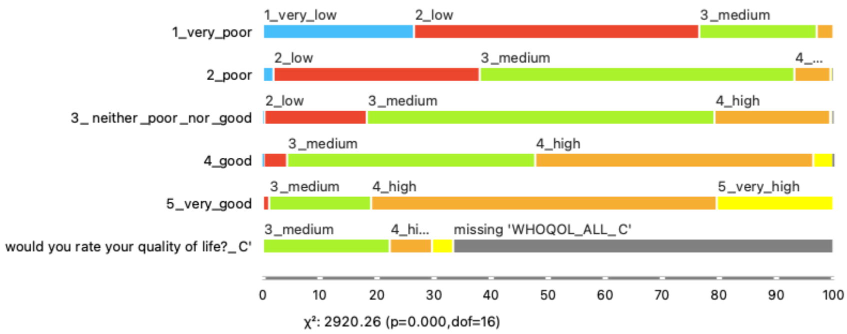

Figure 6. Correlation between the self-assessment question (categorical) and the average of the domain score (categorical).

4.4. Models

Changing the target to q76 (19), different model types were learned. Table 3 shows the performance of each model type. AUC is the "Area Under Curve"; CA is the Classification Accuracy; F1 is the weighted harmonic mean of precision and recall; LogLoss is the cross-entropy loss that takes into account the uncertainty of your prediction based on how much it varies from the actual label [4].

4.5. Hyperparameter Search

As outlined in the previous section, the best-performing model type was the logistic regression. Therefore, a hyperparameter search was done to try and find the optimum values. Figure 7 shows the change in the evaluation parameters regarding the regularization strength.

As figure 7 shows, the model's behaviour appears continuous, and the evaluation parameters seem to converge for regularization strengths greater than 10 for this database and the features selected.

Table 2. Best correlations to the overall average score between the WHOQoL-Bref domains.

Correlation of Average WHOQoL Score	Pearson	Spearman
Psychological	0.850	0.859
Physical	0.768	0.783
Social	0.753	0.769
Environmental	0.740	0.759
q76 (19) How satisfied with yourself?	0.732	0.742
q77 (20) How satisfied are you with your personal relationships?	0.670	0.683
q74 (17) How satisfied are you with your ability to perform your daily living activities?	0.637	0.653
q67 (10) Do you have enough energy for everyday life?	0.625	0.637
q62 (5) How much do you enjoy life?	0.623	0.634

Table 3. Results of different models for q76(19) as target and other features as inputs.

Model	AUC	CA	F1	Precision	Recall	Specificity
Naive Bayes	0.822	0.552	0.556	0.575	0.552	0.844
Tree	0.742	0.576	0.576	0.578	0.576	0.813
SVM	0.854	0.617	0.614	0.617	0.617	0.819
Random Forest	0.844	0.634	0.629	0.633	0.634	0.822
kNN	0.853	0.647	0.638	0.653	0.647	0.820
Gradient Boosting	0.889	0.701	0.696	0.698	0.701	0.856
Logistic Regression	0.994	0.964	0.964	0.964	0.964	0.988

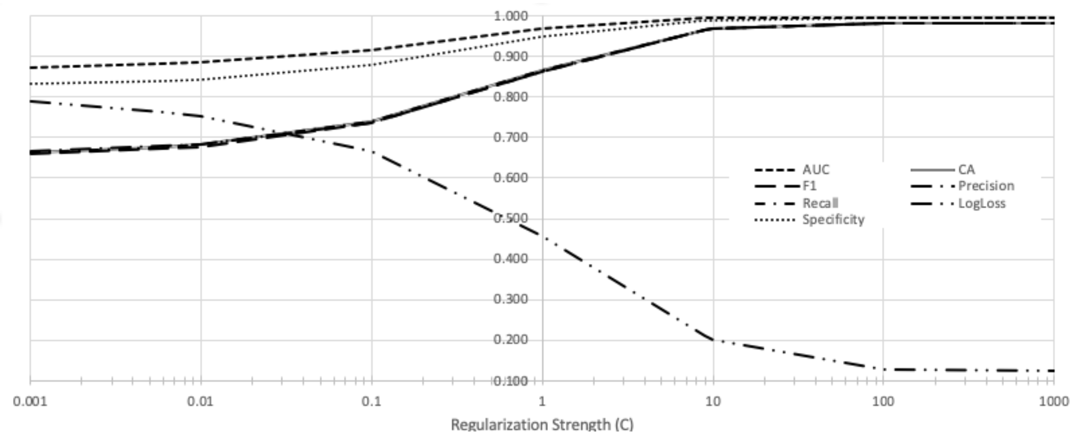

Figure 7. Logistic regression for q76(19) as target in regard to regularization strength (c)

Figure 8 shows the confusion matrix for the Logistic regression model with c=10 in terms of the number of instances.

Table 4. Logistic regression results for q 76(19) as the target in regard to regularization strength (C)

Logistic Regression	Original	1	**2**	3	4	5
Regularization Type	Ridge (L2)	Ridge (L2)	**Ridge (L2)**	Ridge (L2)	Ridge (L2)	Ridge (L2)
Regularization Strength (C)	0.006	100	**10**	1	0.1	0.01
Balance class distribution	NO	NO	**NO**	NO	NO	NO
AUC	0.880	0.995	**0.994**	0.969	0.916	0.884
CA	0.677	0.981	**0.967**	0.866	0.741	0.681
F1	0.672	0.981	**0.967**	0.862	0.735	0.676
Precision	0.678	0.981	**0.967**	0.861	0.737	0.681
Recall	0.677	0.981	**0.967**	0.866	0.741	0.681
LogLoss	0.763	0.128	**0.202**	0.458	0.666	0.753
Specificity	0.839	0.994	**0.989**	0.947	0.879	0.843

Predicted

Actual	1_very_dissatisfied	2_dissatisfied	3_neither_satisfied_nor_dissatisfied	4_satisfied	5_very_satisfied	Σ
1_very_dissatisfied	309	73	0	0	1	383
2_dissatisfied	25	1316	25	0	0	1366
3_neither_satisfied_nor_dissatisfied	0	14	2908	8	2	2932
4_satisfied	1	1	15	1733	17	1767
5_very_satisfied	0	1	2	61	350	414
Σ	335	1405	2950	1802	370	6862

Figure 8. Confusion matrix for the Logistic Regression model (c=10) - total number of instances)

5. Discussion

The esteem question correlates better with the average domain score than the self-assessment question. This is interesting, for it is not expected. Further analysis is necessary to evaluate the reason for this discrepancy.

It was possible to construct a precise and accurate model to predict the esteem question, and the model improved with higher regularization strength on the logistic regression topology. Interestingly, there seems to be a saturation on the regularization strength above $C = 10$. The performance in regards to the regularization strength seems to be continuous, which would allow a gradient search on an eventual automatic ML.

A question for future works is the minimum parameter set for this model and what features are more critical for the prediction.

6. Conclusion

Results show that it is possible to create models to predict some answers to the WHOQOL-Bref questionnaire based on the domain scores and the other questions, at least for the given database.

For this database and model, the evaluation parameters that predict the target features behaved in a way that allows the possibility of using an automatic machine learning algorithm to create optimized models to predict other targets.

7. Limitations

All results and conclusions are limited to this databank and its data. The questionnaires were filled in a fixed time frame. Currently, it is still not possible to infer causality between features.

Finally, with the data at hand, one cannot infer results on other similar datasets without further investigations.

8. Next Steps

The model pipeline had been taken from a reference about automatic model creation. One opportunity is to investigate the possibility of creating an automated learning model framework for problems based on this database and its applicability to similar or dissimilar data samples. Also, simplifying the model by reducing the number of features may provide more generalizable results.

Also, the data exploration was confined to the WHOQOL-Bref part of the data. There is interest in correlating this data with other features present on the questionnaire and assessing the possibility of using the WHOQOL-Bref as a proxy or other targets outside QoL.

9. Conflicts of Interest

The authors do not recognize any conflict of interest related to this research.

10. Acknowledgements

The authors would like to acknowledge CNPq - Conselho Nacional de Desenvolvimento Científico e Tecnológico (164093/2021-4) - and FAPESP - Fundaçao de Amparo à Pesquisa do Estado de São Paulo (2017/01842-6) for the continuing financial support

References

[1] Amenomori, C., Mizumoto, T., Suwa, H., Arakawa, Y., Yasumoto, K.: A method for simplified HRQOL measurement by smart devices. Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering pp. 91–98 (2018)

[2] Batata, O., Augusto, V., Xie, X.: Caregivers burnout prediction using supervised learning. 2018 IEEE International Conference on Systems, Man, and Cybernetics (SMC) (oct 2018). https://doi.org/https://doi.org/10.1109/SMC.2018.00302

[3] Demenech, L.M., Oliveira, A.T., Neiva-Silva, L., Dumith, S.C.: Prevalence of anxiety, depression and suicidal behaviors among brazilian undergraduate students: A systematic review and meta-analysis. Journal of Affective Disorders **282**, 147–159 (mar 2021). https://doi.org/https://doi.org/10.1016/j.jad.2020.12.108

[4] Demšar, J., Curk, T., Erjavec, A., Črt Gorup, Hočevar, T., Milutinovič, M., Možina, M., Polajnar, M., Toplak, M., Starič, A., Štajdohar, M., Umek, L., Žagar, L., Žbontar, J., Žitnik, M., Zupan, B.: Orange: Data mining toolbox in python. Journal of Machine Learning Research **14**, 2349–2353 (2013), `http://jmlr.org/papers/v14/demsar13a.html`

[5] Division of Mental Health and Prevention of Substance Abuse - World Health Organization: WHOQOL User Manual (1998)

[6] Goecks, J., Jalili, V., Heiser, L.M., Gray, J.W.: How machine learning will transform biomedicine. Cell **181**(1), 92–101 (apr 2020). https://doi.org/https://doi.org/10.1016/j.cell.2020.03.022

[7] Juliano, I., , dos Santos Junior and, A.: Perfil descritivo dos estudantes de graduação na área da saúde da unicamp no ano de 2017 quanto a sexualidade, comportamento sexual, orientação sexual e identidade de gênero. In: Revista dos Trabalhos de Iniciação Científica da UNICAMP. Universidade Estadual de Campinas (nov 2019). https://doi.org/DOI: 10.20396/revpibic2720191680

[8] Kaur, M., Dhalaria, M., Sharma, P.K., Park, J.H.: Supervised machine-learning predictive analytics for national quality of life scoring. Applied Sciences **9**(8), 1613 (apr 2019). https://doi.org/https://doi.org/10.3390/app9081613

[9] Lee, S., Chung, J.Y.: The machine learning-based dropout early warning system for improving the performance of dropout prediction. Applied Sciences **9**(15), 3093 (jul 2019). https://doi.org/doi:10.3390/app9153093

[10] Macalli, M., Navarro, M., Orri, M., Tournier, M., Thiébaut, R., Côté, S.M., Tzourio, C.: A machine learning approach for predicting suicidal thoughts and behaviours among college students. Scientific Reports **11**(1) (jun 2021). https://doi.org/10.1038/s41598-021-90728-z

[11] Moutinho, I.L.D., Lucchetti, A.L.G., da Silva Ezequiel, O., Lucchetti, G.: Mental health and quality of life of brazilian medical students: Incidence, prevalence, and associated factors within two years of follow-up. Psychiatry Research **274**, 306–312 (Apr 2019). https://doi.org/https://doi.org/10.1016/j.psychres.2019.02.041

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

[13] Pinto, M., Marotta, N., Caracò, C., Simeone, E., Ammendolia, A., de Sire, A.: Quality of life predictors in patients with melanoma: A machine learning approach. Frontiers in Oncology **12** (mar 2022). https://doi.org/https://doi.org/10.3389/fonc.2022.843611

[14] Quagliato, I., , dos Santos Junior and, A.: O estudante do PROFIS: perfil sócio-demográfico, cultural, identidade pessoal e social, espiritualidade, sexualidade, qualidade de vida, uso de álcool e outras substâncias psicoativas, saúde física e mental. In: Revista dos Trabalhos de Iniciação Científica da UNICAMP. Universidade Estadual de Campinas (nov 2019). https://doi.org/https://doi.org/10.20396/revpibic2720192493

[15] Sekeroglu, B., Dimililer, K., Tuncal, K.: Student performance prediction and classification using machine learning algorithms. Proceedings of the 2019 8th International Conference on Educational and Information Technology (mar 2019). https://doi.org/http://doi.org/10.1145/3318396.3318419

[16] Srividya, M., Mohanavalli, S., Bhalaji, N.: Behavioral modeling for mental health using machine learning algorithms. Journal of Medical Systems **42**(5) (apr 2018). https://doi.org/https://doi.org/10.1007/s10916-018-0934-5

[17] THE WHOQOL GROUP: Development of the world health organization WHOQOL-BREF quality of life assessment. Psychological Medicine **28**(3), 551–558 (may 1998). https://doi.org/https://doi.org/10.1017/S0033291798006667

[18] Zoller, M.A., Huber, M.F.: Benchmark and survey of automated machine learning frameworks. Journal of Artificial Intelligence Research 70 (2021) 409-472 (Apr 2019)