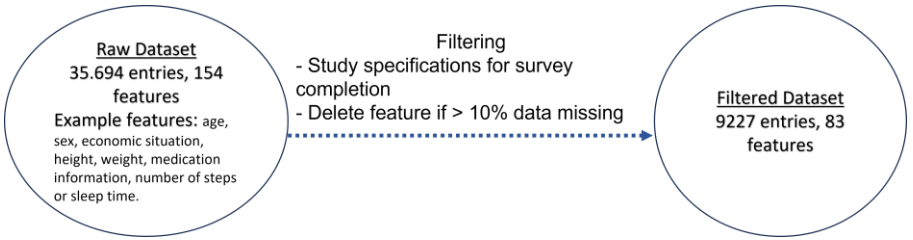
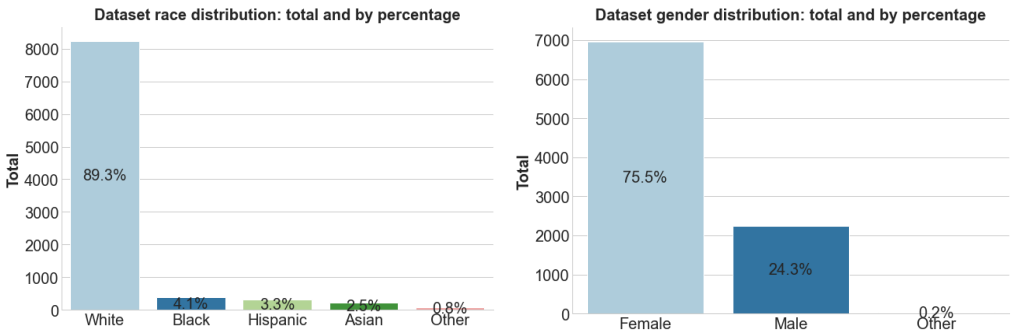
Machine Learning Project: Depression severity prediction

This Machine Learning Project consists of predicting the depression severity category (from 0 to 4) from self-reported health data and wearable data collected in a 1-year span. The model selection and hyperparameter tuning were based on optimizing the weighted F1-score. Results can be compared to those obtained by the paper [1] and are shown on the results section.

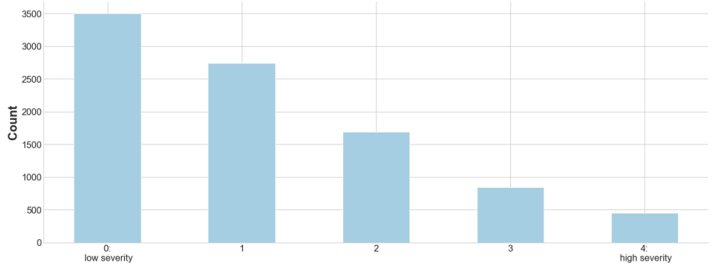
Dataset



The dataset's sampling is heavily biased towards white and female individuals:



Multiclass and imbalanced target (self-reported depression severity) for classification



Feature Engineering

Counting
Sum all negative characteristics (booleans) such as:
Individual has a chronic illness (diabetes type 1 or 2, gout...), experienced injury last year, started a new medication...

Grouping
Assign the mean/median of a feature by group (gender, race, education level...) to each entry.
For example, calculate the median BMI by gender
`df.groupby("sex")["bmi"].transform("median")`

Interactions and polynomial transforms
Squared or cubed feature values
Multiply values from different features together

Thresholds
Walked more than 8000 steps in average
Slept between 300 and 500 minutes on average

[1]: Makhmutova, A. et al.(2021). "Prediction of self-reported depression scores using Person-Generated Health Data from a Virtual 1-Year Mental Health Observational Study" Proceedings of the 2021 Workshop on Future of Digital Biomarkers, Pages 4–11.

Preprocessing

Feature preprocessing consists of imputation when values are missing (most frequent and median strategies), scaling (by removing the mean and scaling to unit variance), normalization, binning and one-hot encoding.

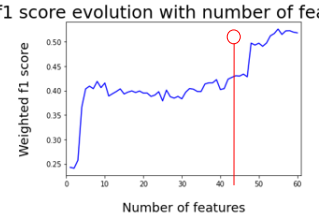
Evaluation

Metrics: For comparison with [1], the weighted f1, weighted quadratic Cohen-Kappa and adjacent accuracy scores are selected.
Nested cross-validation: For generalization evaluation and hyperparameter tuning. Results for the full dataset with the 50 additional created features:

Model	Weighted F1	Cohen-Kappa	Adjacent Accuracy
Logistic Regression	0.426	0.437	0.832
HistGradientBoostingClassifier	0.431	0.432	0.830
XGBoost	0.428	0.421	0.830

XGBoost Feature Selection

Evaluate model by adding feature by feature. 57 features are chosen.



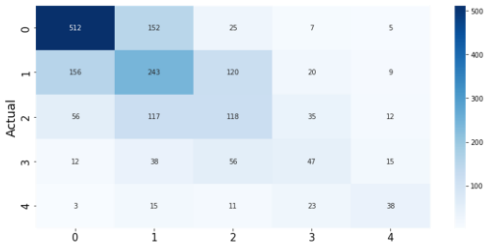
Iterative process

XGBoost Hyperparameter tuning

Use of hyperopt library, tuned 12 parameters for best weighted f1-score

Final Model Results

Metric	Adjac. Acc.	Balanced Acc.	Cohen-Kappa	Weighted F1
Results	0.89	0.52	0.62	0.52
Paper	0.89	0.47	0.66	0.54



To Be Done

Error analysis, SMOTE