SAMSUNG

Samsung Innovation Campus

AI Course

Together for Tomorrow!
Enabling People
Education for Future Generations

Smart Stress Detection

AI Course

Problem Definition

Smart Stress Detection

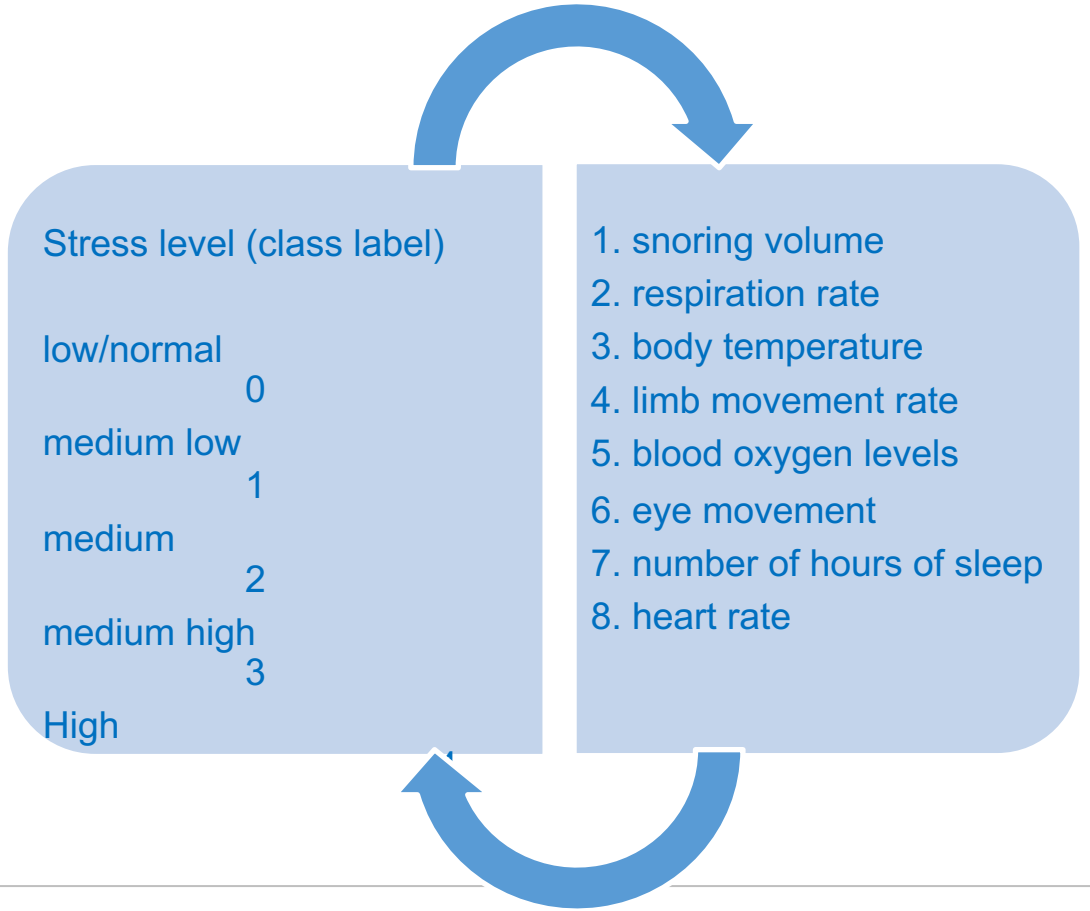
- Many health problems related to stress like obesity, heart disease, Alzheimer's disease, diabetes, depression, and gastrointestinal problems

The Aim → predict the level of the stress using ML algorithms based the sleeping characteristics.

Data Source and Data Description

Human Stress Detection in and through Sleep (SaYoPillow) from Kaggle.com

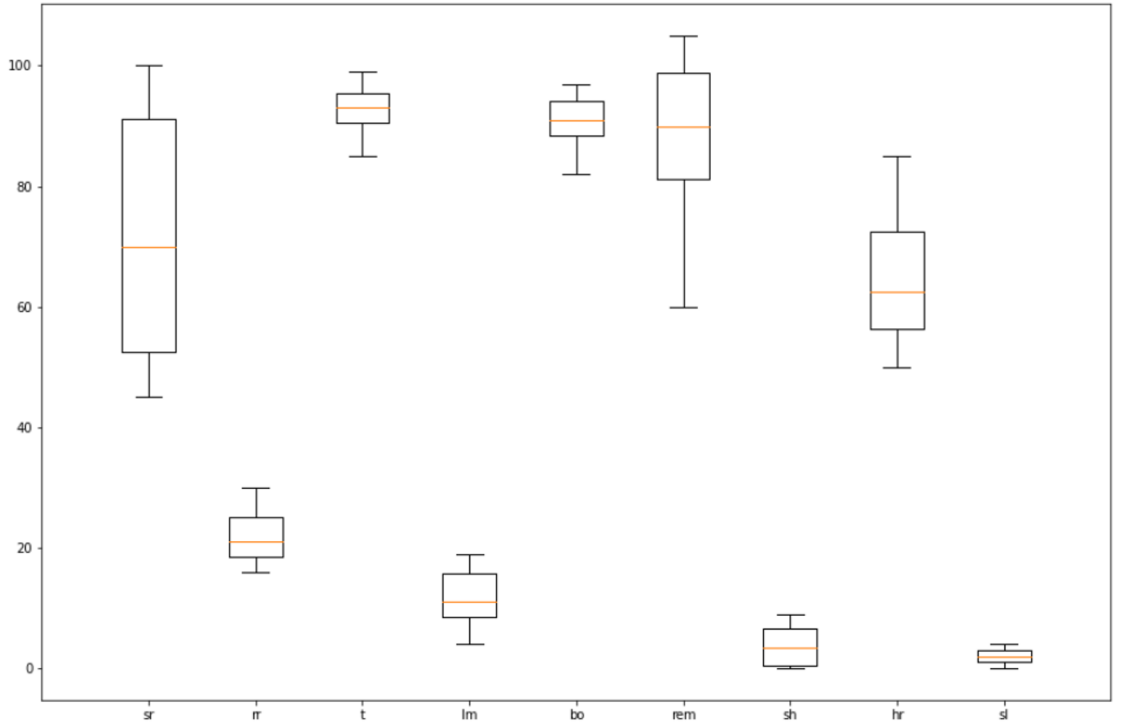
The relationship



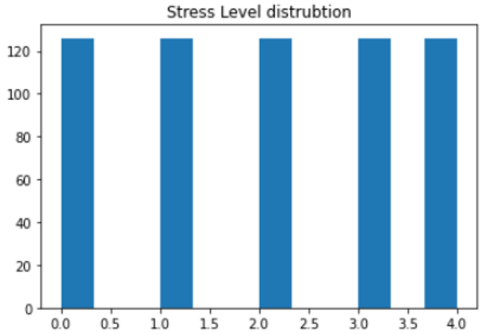
I. Data Preprocessing

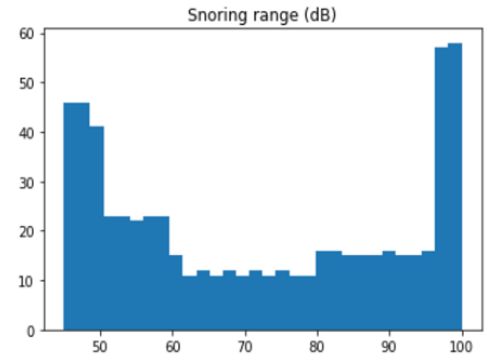
Outliers and null values

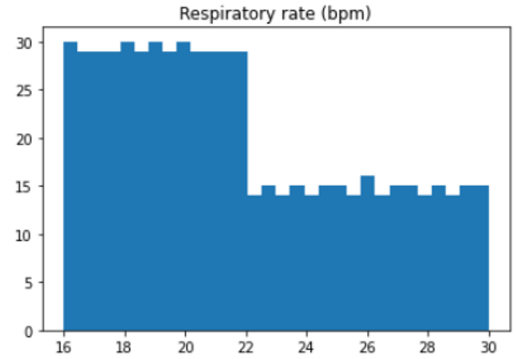
```
RangeIndex: 630 entries, 0 to 629
Data columns (total 9 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   sr      630 non-null    float64
 1   rr      630 non-null    float64
 2   t       630 non-null    float64
 3   lm      630 non-null    float64
 4   bo      630 non-null    float64
 5   rem     630 non-null    float64
 6   sh      630 non-null    float64
 7   hr      630 non-null    float64
 8   sl      630 non-null    int64
```

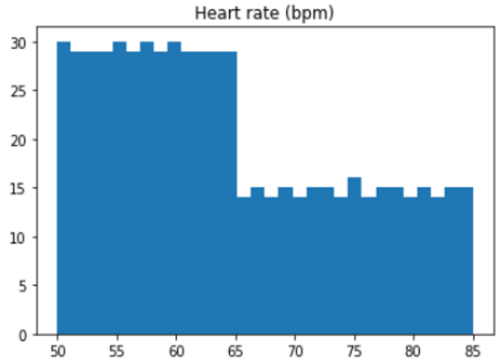


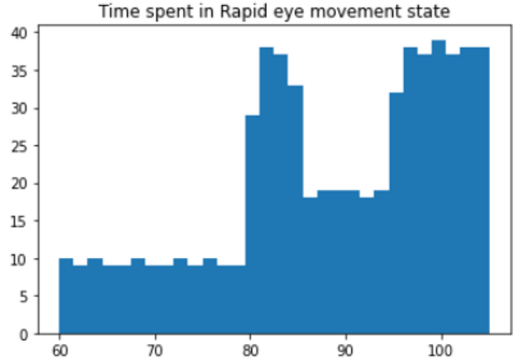
Distributions

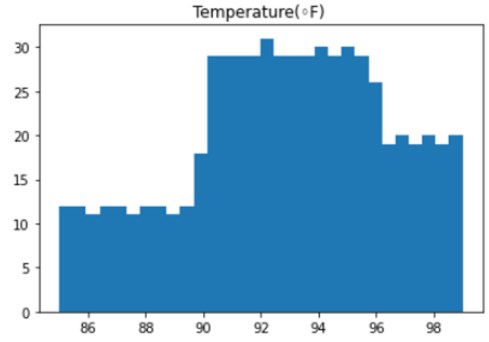


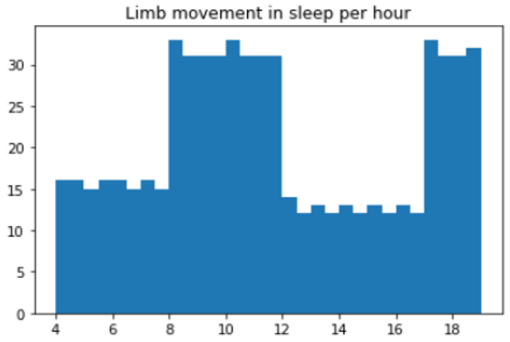


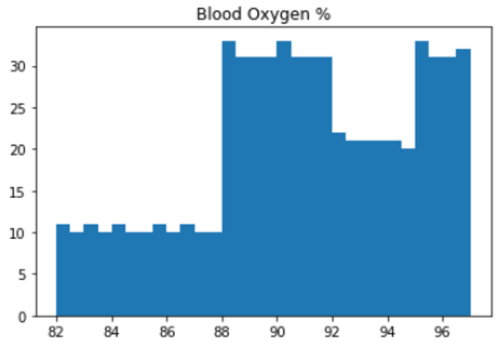


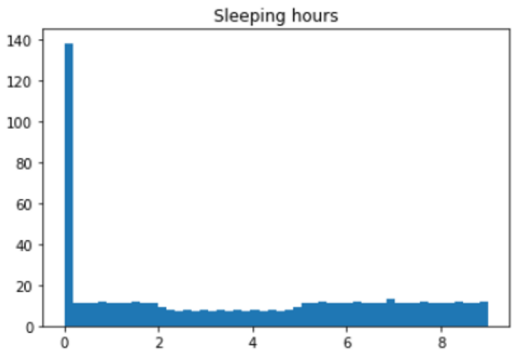




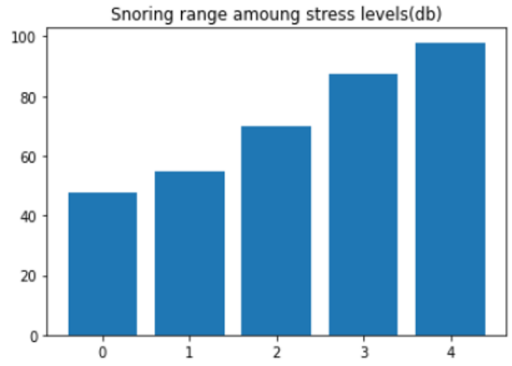


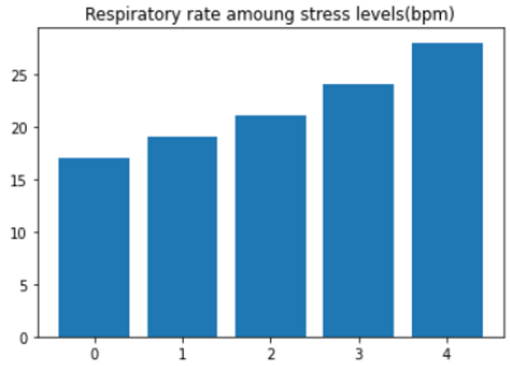


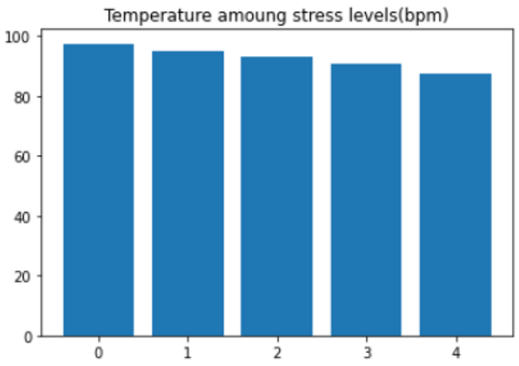


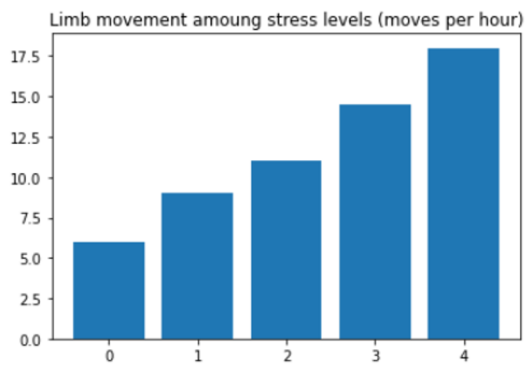


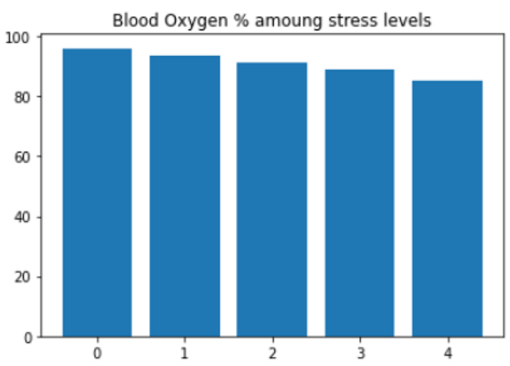
Means according to stress levels

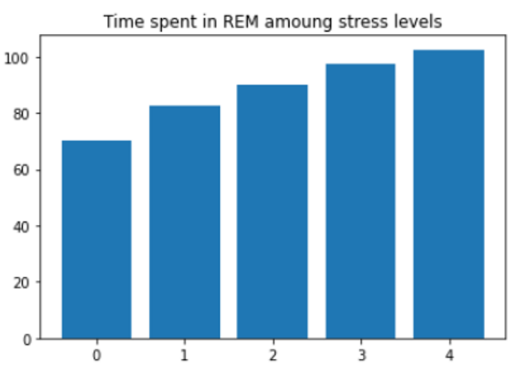


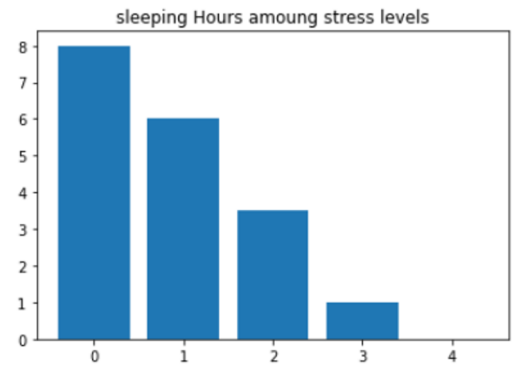


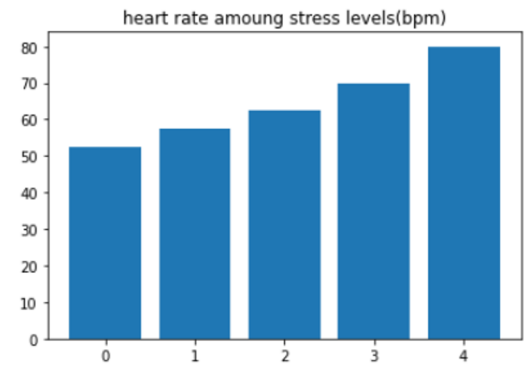












ANOVA

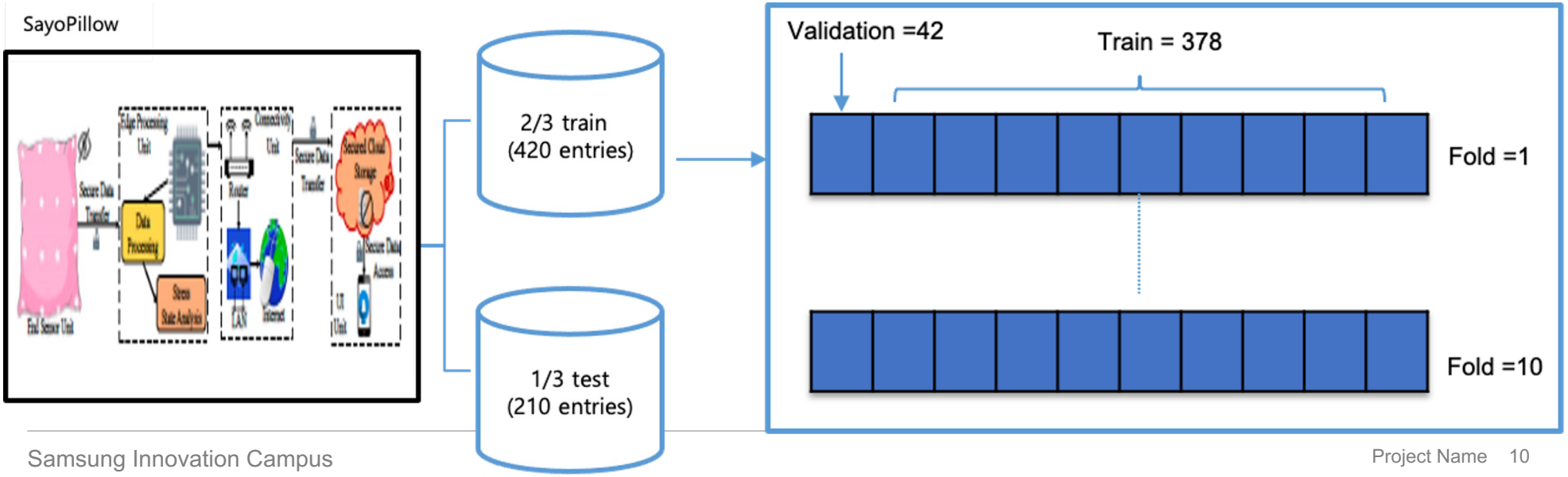
- Snoring range p value < 0.005
- Respiratory rate p value < 0.005
- Temperature p value < 0.005
- Limb movement p value < 0.005
- Blood Oxygen p value < 0.005
- REM p value < 0.005
- Sleeping Hours p value < 0.005
- Heart rate p value < 0.005

4. Scaling and Splitting Data

Z-score normalization

	sr	rr	t	lm	bo	rem	sh	hr	sl
0	0.887273	0.691429	0.488571	0.840000	0.522667	0.880000	0.204444	0.691429	3
1	0.848000	0.650286	0.468000	0.792000	0.503467	0.864000	0.172444	0.650286	3
2	0.272727	0.285714	0.785714	0.400000	0.866667	0.555556	0.777778	0.285714	1
3	0.741091	0.538286	0.412000	0.661333	0.451200	0.820444	0.085333	0.538286	3
4	0.056727	0.089143	0.919429	0.166400	0.949867	0.277333	0.916444	0.089143	0

2/3 for train and 1/3 for test

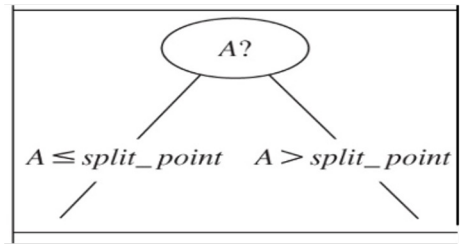


II. Machine Learning Modeling

CART decision tree algorithm

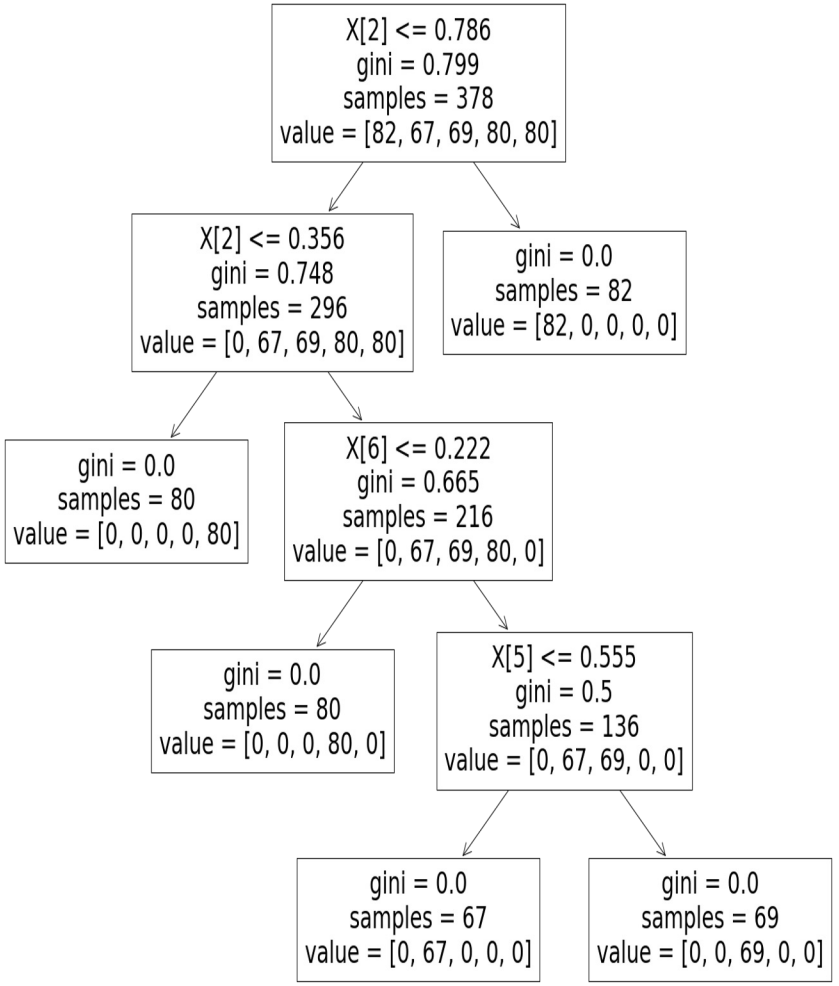
Classification And Regression Trees (CART) algorithm

- Binary Tree



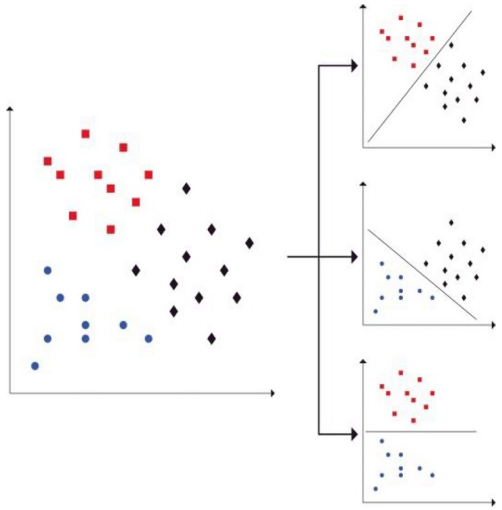
- Gini's impurity index to select attribute

$$Gini(D) = 1 - \sum_{i=1}^{m} p_i^2,$$

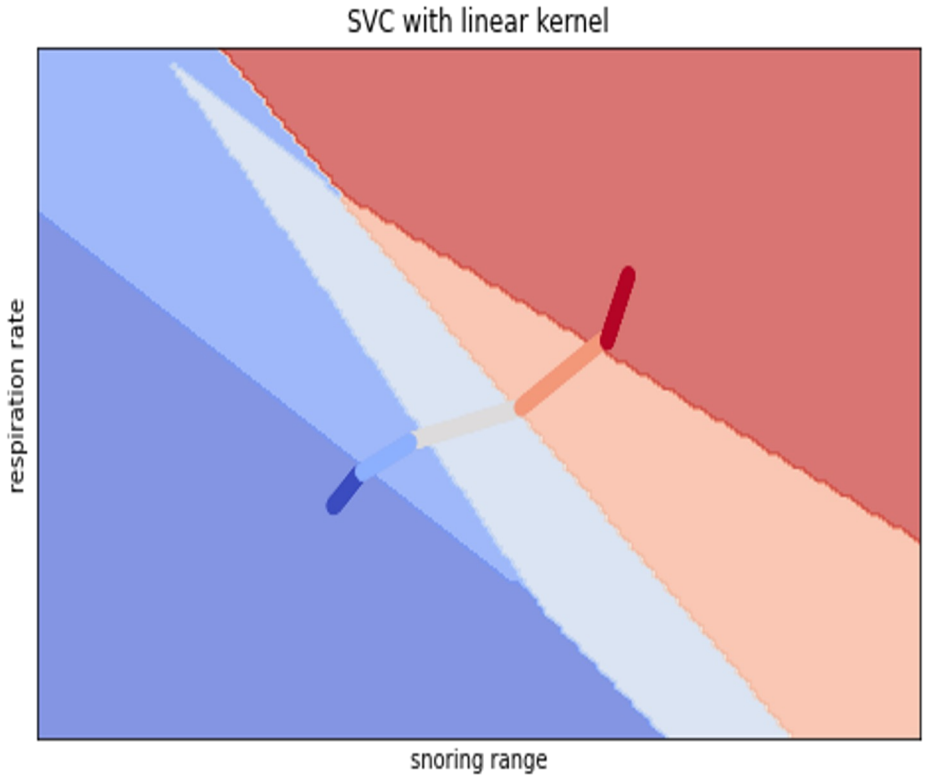


Linear kernel with one-versus-one approach for multiclass

Split the dataset into one dataset for each class opposite to every other class.



→The majority counts



XGBoost Algorithm

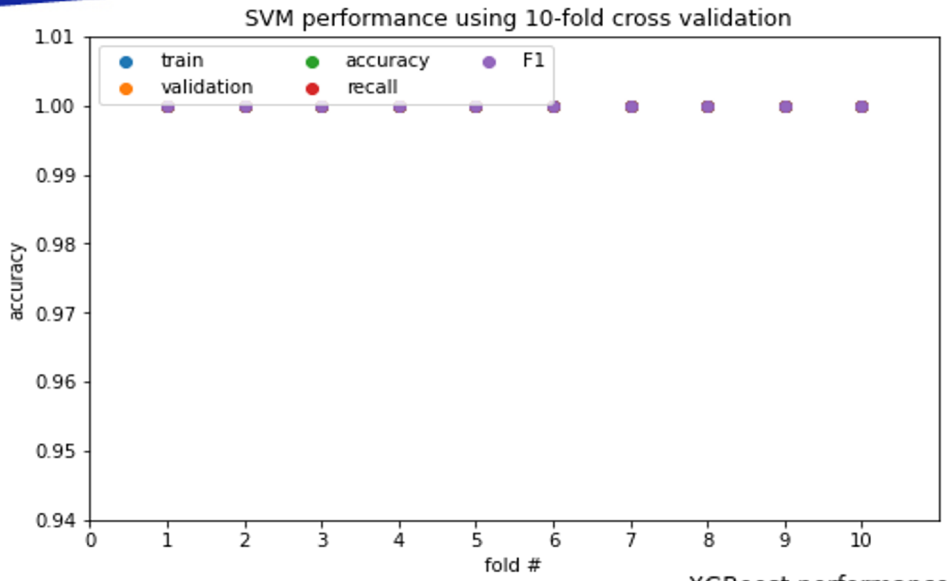
- Gradient boosted decision trees
- Push the limit of computations resources for boosted tree algorithms
- Ensemble technique
- Gradient descent algorithm to minimize the loss

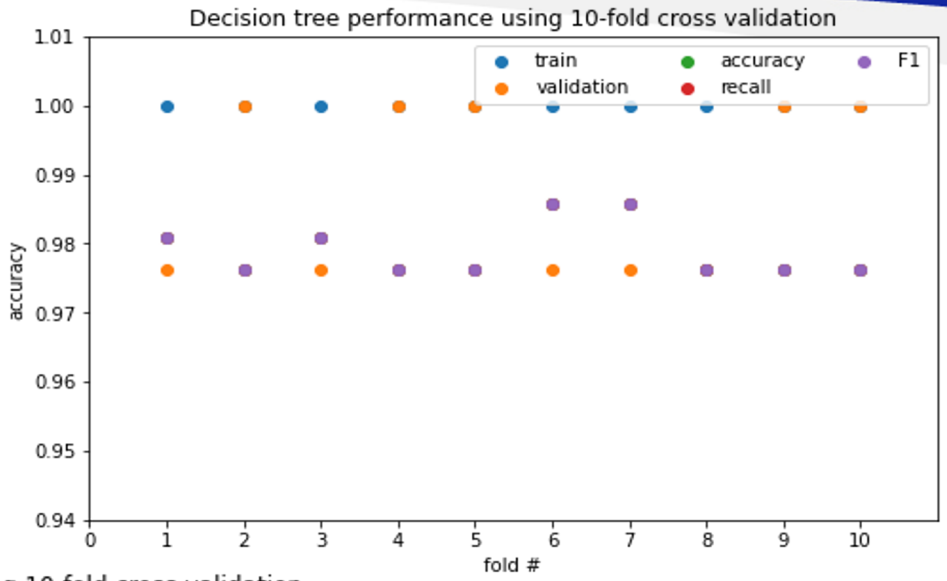
III. Performance metrics

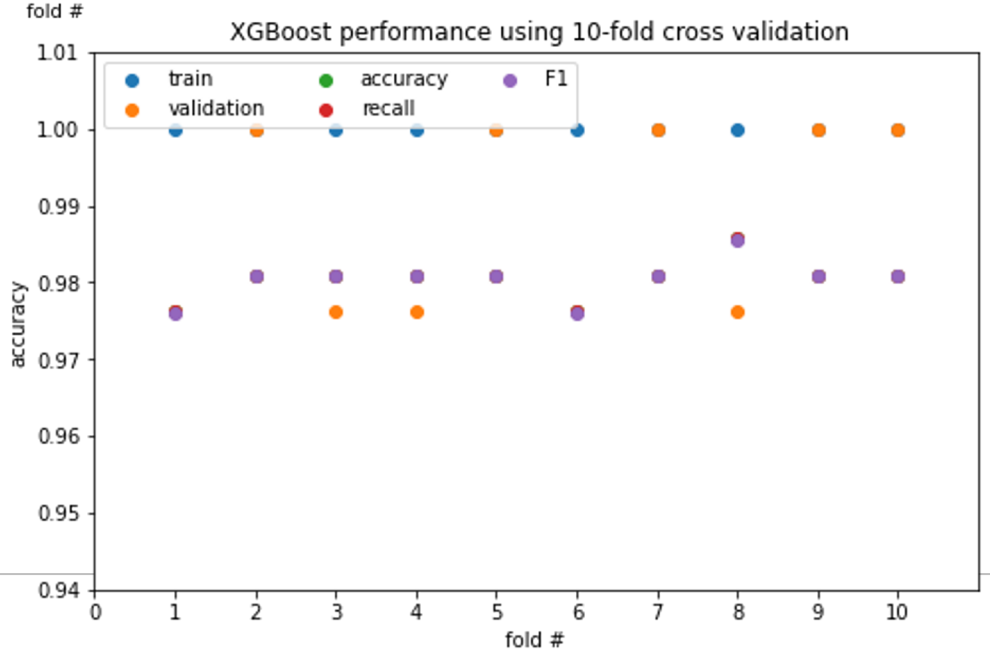
Overall Performance

Metrics\Model	Decision Tree	SVM	XGBoost
Accuracy	97.6%	100%	98.09%
Recall	97.6%	100%	98.09%
F1	97.6%	100%	98.08%

Overall Performance







Conclusion

We believe that the dataset is clean and too simple to be modeled as machine learning task.

Snoring Range (dB)	Respiration rate (bpm)	Heart rate (bpm)	Stress State
50-80 dB	17-22 bpm	54-64 bpm	Low, Medium Low and Medium Stress State
80-89 dB	23-25 bpm	65-70 bpm	Medium High Stress State
90+	25+	70+	High Stress State