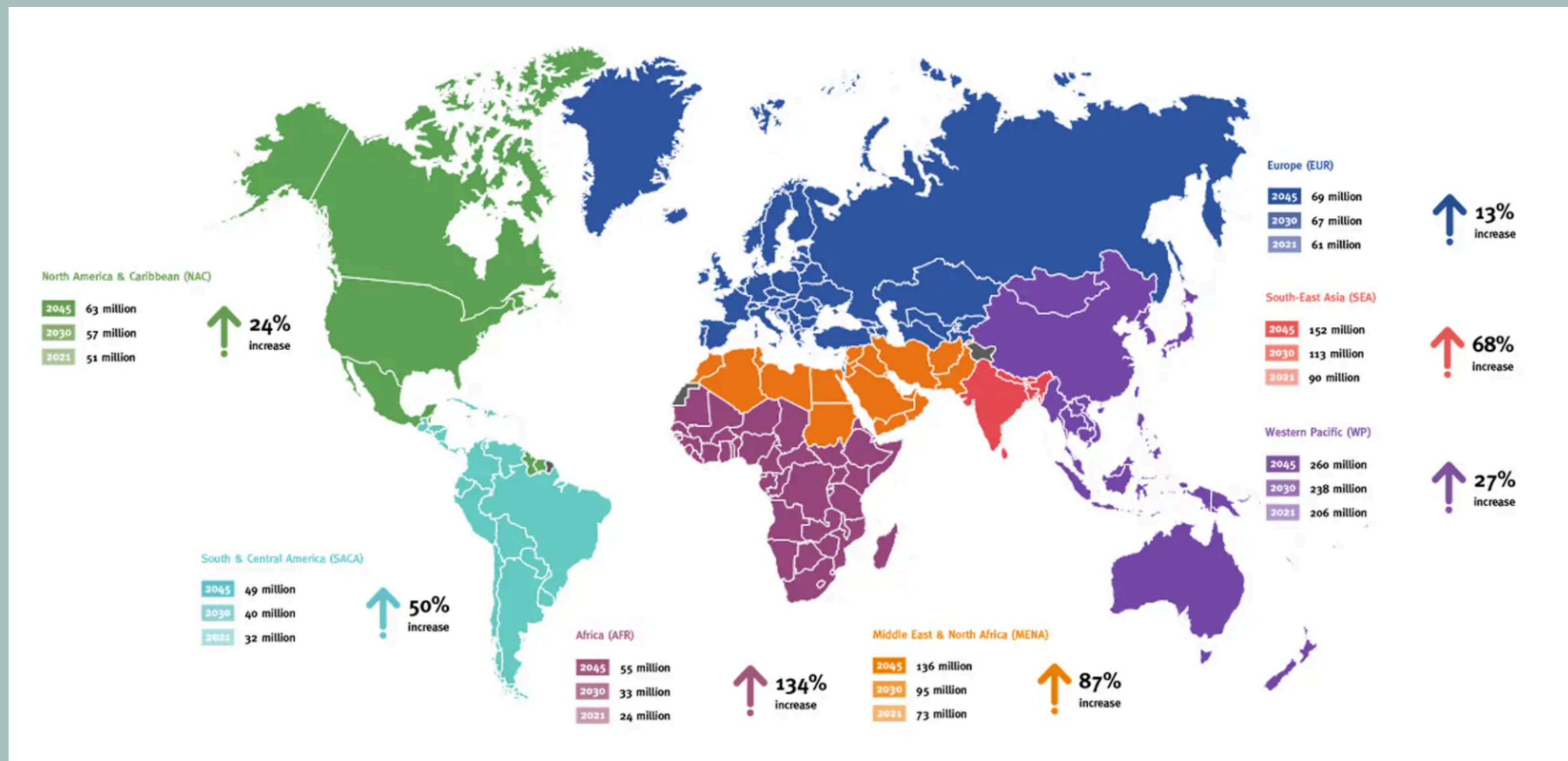




Eduardo Martínez Manzanero





# Contents

- 01. Exploratory Data Analysis
- 02. Data Preprocessing
- 03. ML Model Aproximation
- 04. Model Improvement
- 05. Conclusions



# Target: Multiclass

Non diabetic

Prediabetic

Diabetic

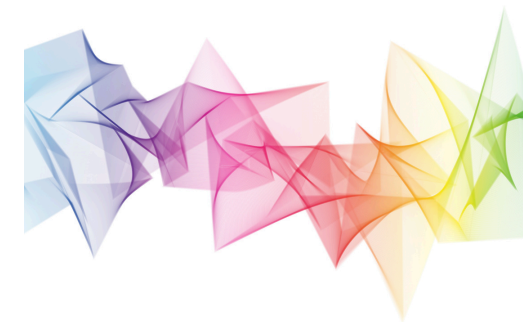


# Model election



## Decision Tree Classifier

0	1
1	0
2	0



## Linear SVC

0	0.99
1	0
2	0.09



## KNN

0	0.99
1	0
2	0.09

# Improving the model

☆ **Binarizing**



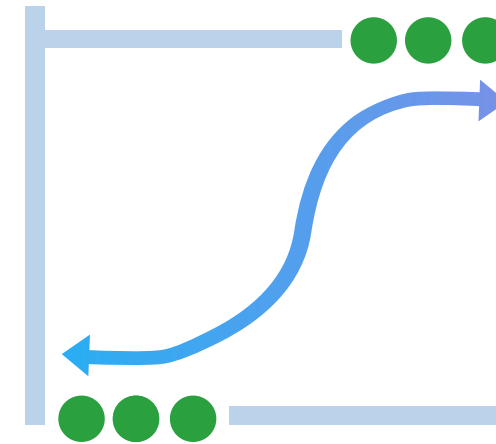
☆ **Downsampling**



☆ **Scaling**



# Model selection



Logistic Regression

0	0.71
1	0.76



Decision Tree Classifier

0	0.71
1	0.71



KNN

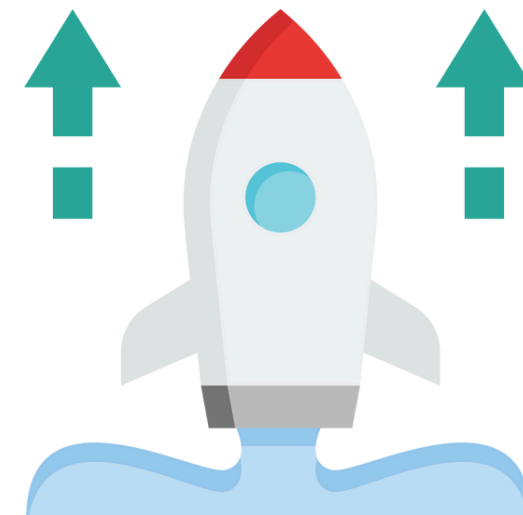
0	0.68
1	0.75

# Model selection



Linear SVC

0	0.72
1	0.77



XGBoost

0	0.72
1	0.77



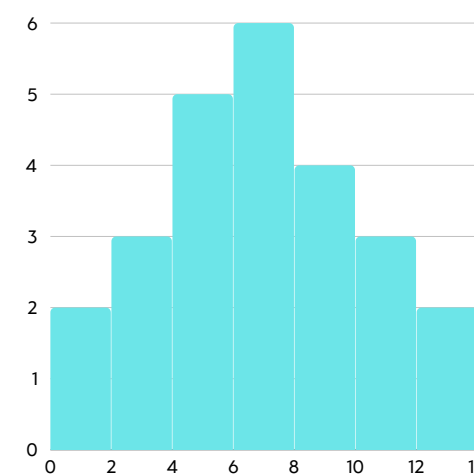
# Scaling Model Selection

Over XGBoost

0

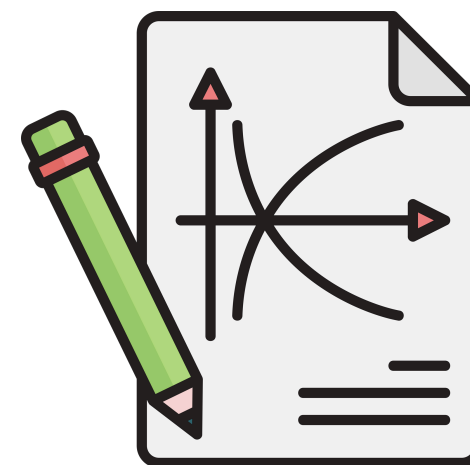
Min- Max Scaler

0	0.71
1	0.78



Standar Scaler

0	0.92
1	0.31



Log Transform

0	0.71
1	0.78

# Dropping columns

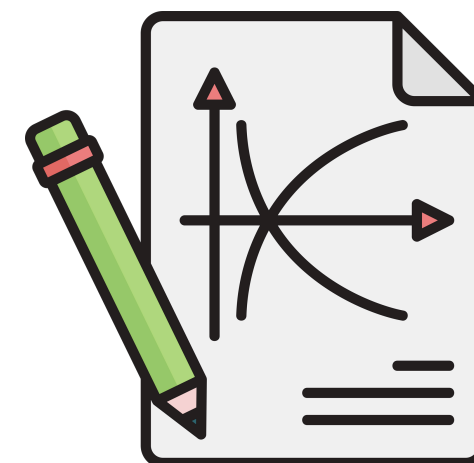


(columns=  
["menthlth", "physhlth", "education",  
"diffwalk", "fruits"])



XGBoost

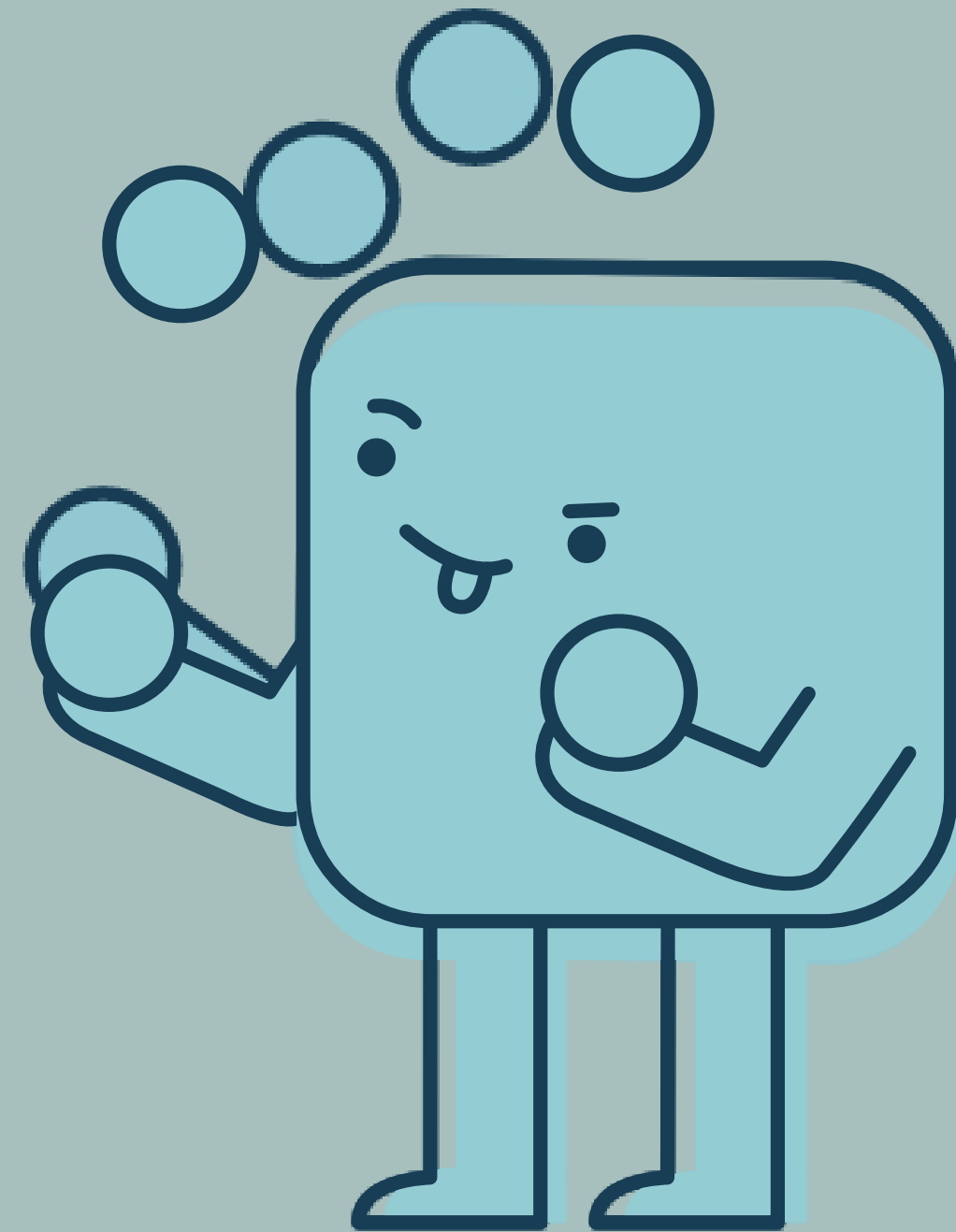
0	0.71
1	0.78



Log Transform

0	0.70
1	0.79

Cross- validation



# Conclusiones

# ¡Muchas gracias!

*Eduardo Martínez Manzanero*