Multi-Class Prediction of Obisity risk

kaggle

Top 1%

Data Science Competition

Power BI

LightGBM

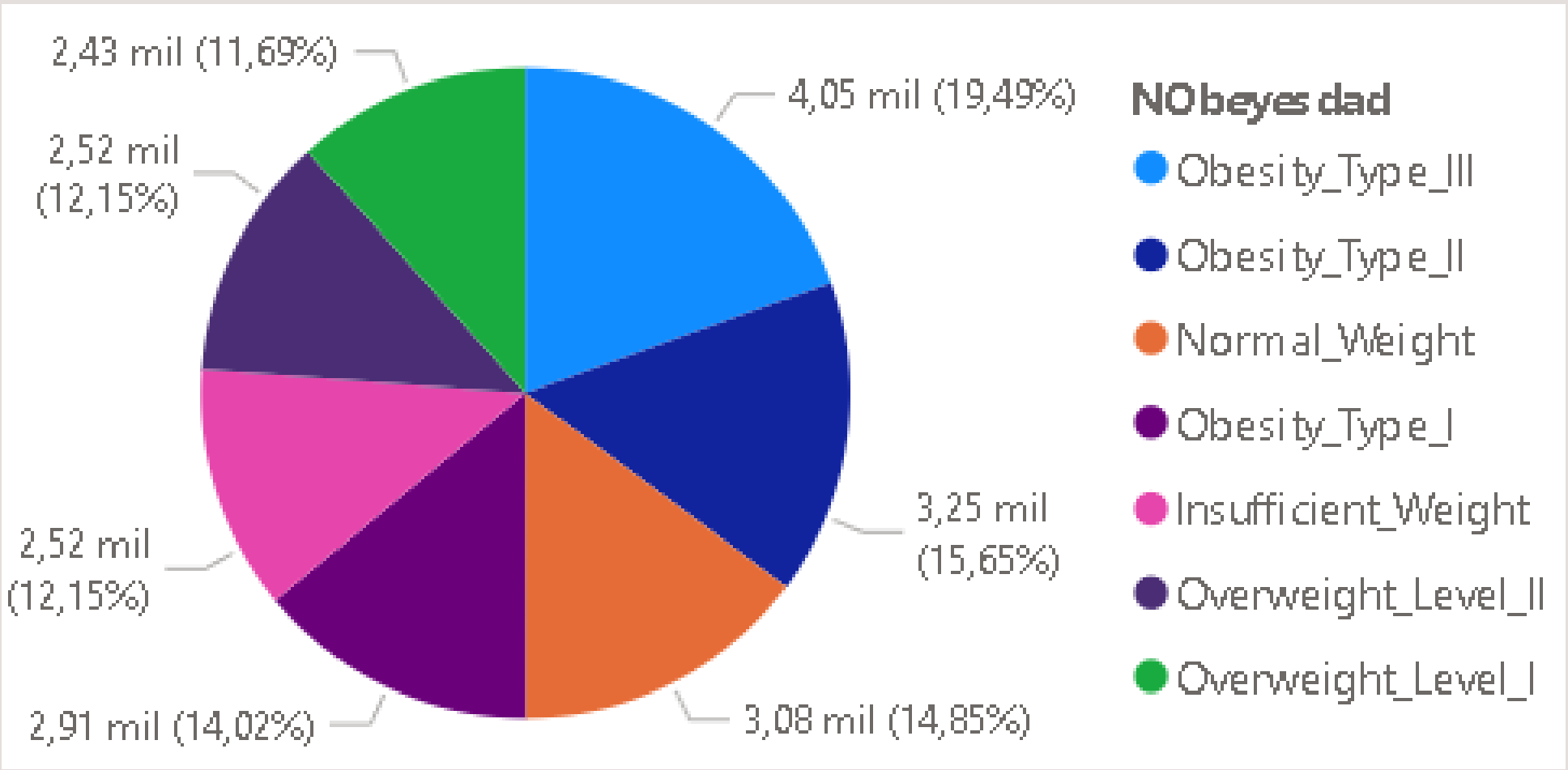
AMALIO GÓMEZ LÓPEZ

Summary

- Exploratory Data Analysis
- Baseline Submission
- Feature engineering
- XGB / LGBM Models (Weighted combination of predictions)
- Hyperparameter Optimization using Optuna
- Feature Importance

EDA Conclusions

- Target variable is quite balanced.
- Age, height, weight are the most relevant. Focus on for feature engineering.
- Differences in distributions E.g. Age



Pie chart illustrating percentages of TARGET to classify

Feature Engineering

•We create Body Mass Index variable
BMI = WEIGHT / (HEIGHT/100) **2

•Interaction between gender and alcohol
Gender _ Alcohol=GENDER * ALCOHOL

•Relationship between family with overweight, Age, Caloric consumption and Number of meals per day

Final Model Combination

In a new DataFrame, create columns to store score and train/test predictions.

We define weights for each model.

Finally, our predictions get combined, we submit using weighted average.

Possible improvement: using a linear model to combine the predictions.

AMALIO GÓMEZ

Thank you for your time

Check the project on GitHub

WWW.AMALIOGOMEZLOPEZ.COM