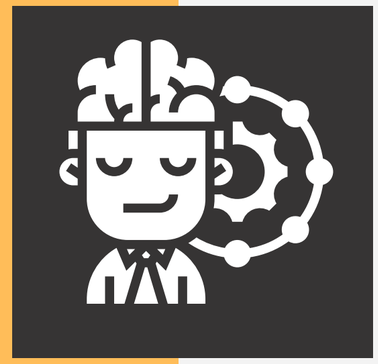
MACHINE LEARNING & ARTIFICIAL INTELLIGENCE



Practical Application Case Study

Group 5:

Federico Barca, Felipe Fischel, Luca Conti, Ayça Basaran, Pablo Ostos

Table of Content

Business Case: Nutrition Business

Nutrition business that helps individuals achieve their target weight. It is essential to have a deep understanding of their health habits and behaviours.

1. **Personalization:** ML enables personalized nutrition plans based on unique health habits.
2. **Accuracy:** ML predicts weight based on health habits with high accuracy.
3. **Efficiency:** ML automates weight prediction process, saving time and resources.
4. **Insights:** ML provides valuable insights into health habits associated with weight gain/loss.

Selecting a dataset

Attribute	Description
Gender	Sex
Age	Age
Weight	Weight
Height	Height
FHWO	Overweight family members
FAVC	Consume high-calorie foods frequently
FCVC	Number of main meals a day
CAEC	Eat food between meals
SMOKE	How often you smoke
CH2O	Liters of water you consume daily
SCC	Monitor calories you consume daily
FAF	Frequency of days per week that you often have physical activity
TUE	Time of use of technological devices on a daily basis
CALC	Frequency of alcohol intake
MTRANS	Means of transportation that you use regularly
NOobeyesdad	Body mass index

Why this dataset?

By analyzing a dataset that includes information on health habits, we can **predict an individual's weight** and develop personalized nutrition plans that are tailored to their needs.

The dataset we have chosen contains **information on individuals'** dietary habits, physical activity levels, and health status. By using machine learning algorithms, we can predict an individual's weight based on these variables with a high degree of accuracy.

Preparing the dataset

1

Changing column names

Original Name	New Name
FAVC	JUNK
FCVC	VeggieCount
NCP	MealCount
CAEC	Snacking
CH20	H20
SCC	CalorieMonitoring
FAF	ActivityFreq
TUE	Screentime
CALC	AlcoholFreq
MTRANS	Transportation
NObeyesdad	BMI

2

Encoding categorical columns

Function Name	Description
encode_yes_no	Converts categorical columns into binary columns.
encode_snack	Returns integer code based on a dictionary mapping.
encode_trans	Returns integer code based on a dictionary mapping.
encode_gender	Converts Gender column to binary column.
encode_bmi	Returns integer code based on a dictionary mapping.
apply()	Applies encoding functions to appropriate columns. Converts to numeric

3

Checking for null values

```
1  obesity_df.isna().sum()
```

```
Gender                            0
Age                               0
Height                            0
Weight                            0
family_history_with_overweight    0
JUNK                              0
VeggieCount                       0
MealCount                         0
Snacking                          0
SMOKE                             0
CH2O                              0
CalorieMonitoring                 0
ActivityFreq                      0
Screentime                        0
AlcoholFreq                       0
Transportation                    0
BMI                               0
```

4

Dropping columns

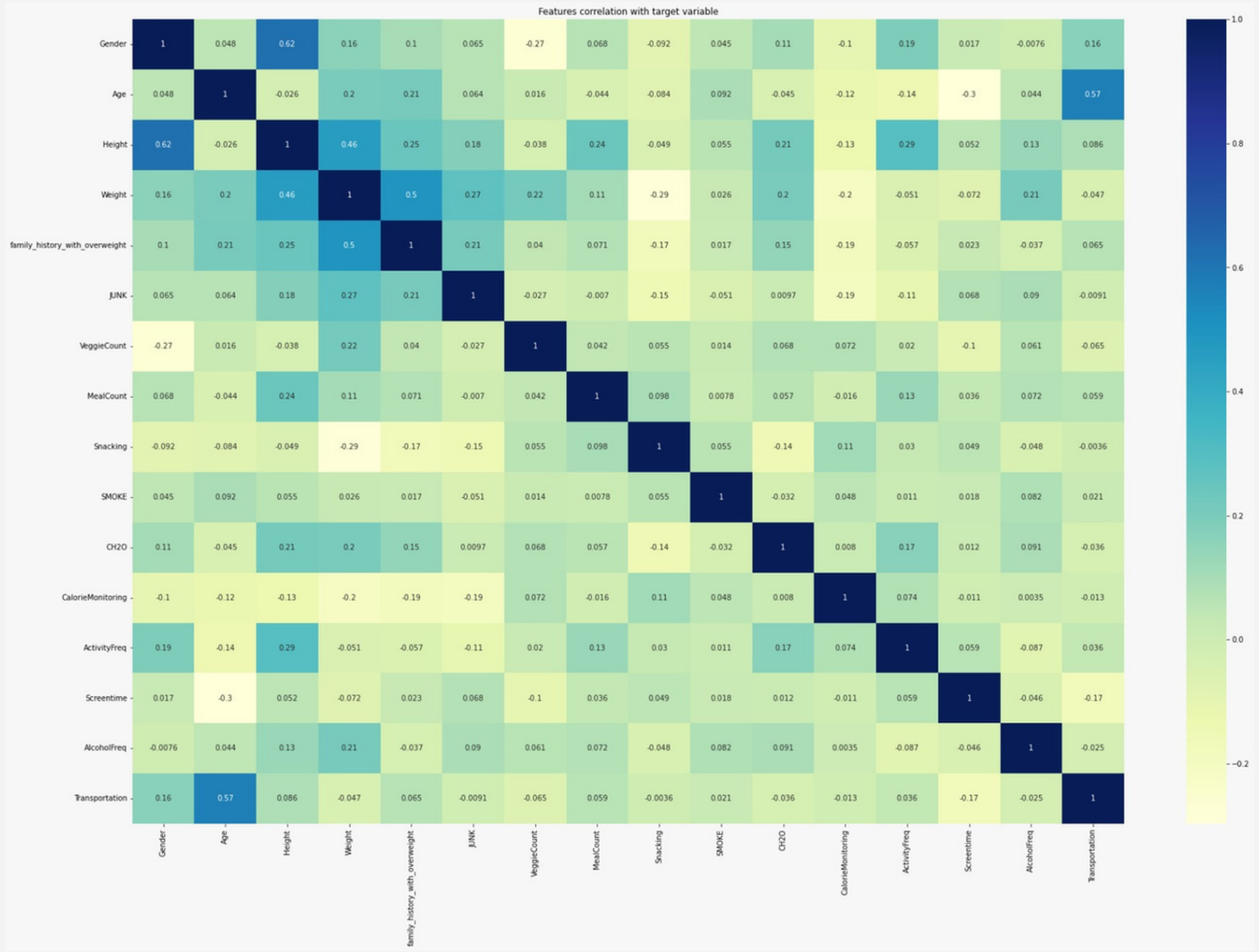
```
obesity_df = obesity_df.drop(columns=['BMI'])
obesity_df
```

BMI

(Body Mass Index)

Correlation matrix

Correlation coefficients between each pair of features in the dataset.



Selecting and Running ML Techniques

Ridge

Regularization techniques that **aim to reduce overfitting** in linear models by adding a penalty term to the cost function.

Random Forest

A tree-based ensemble model that **combines multiple decision trees** to make predictions.

1	2	3	4	5

Linear regression

A simple linear model that predicts a **continuous outcome based on a linear relationship** between the outcome and one or more predictors.

Lasso

Regularization techniques that aim to reduce overfitting in linear models by adding a penalty term to the cost function.

SVR

A type of SVM that predicts a continuous outcome by finding the best hyperplane that maximizes the margin between the predicted values and the actual values.

Comparative analysis

Model	MAE	MSE	RMSE	R2 Score	RMSE (Cross-Validation)
RandomForestRegressor	5.262574	74.706264	8.643279	0.89405	16.356488
SVR	9.43287	175.509946	13.248017	0.751089	24.24906
Ridge	14.106736	310.850615	17.630956	0.559147	24.521576
LinearRegression	14.166991	310.877068	17.631706	0.559109	24.621349
Lasso	16.393615	413.850997	20.343328	0.41307	26.452884

- Best performing: Random Forest Regressor with an RMSE of 8.64 and an R2 score of 0.89. This indicates that the model can explain 89% of the variance in the target variable.
- The SVR comes in second place with an RMSE of 13.25 and an R2 score of 0.75. It performs better than both Ridge and Lasso regression models.
- The Ridge and Linear Regression models perform similarly with an RMSE of 17.63, but Ridge regression performs slightly better on the cross-validation, indicating better generalization.
- The Lasso regression model performs the worst among all models with an RMSE of 20.34 and an R2 score of 0.41.

Informed reflection

Predict Weight

By using machine learning algorithms, we can **predict an individual's weight** based on these variables with a high degree of accuracy.

Personalized plans

Using this predictive model, we can develop **customized nutrition plans** that take into account an individual's unique health habits and preferences.

Health trends

Identify patterns and **trends in health** habits that are associated with weight gain or loss. This information can be used to inform our marketing and outreach strategies

Business

Our nutrition business can deliver personalized and **effective weight management solutions** that help our clients achieve their health goals.

Better Resource Allocation

Allocate resources more **effectively**, such as staff members to clients based on their needs.

Improved Health Outcomes

Improve weight management programs, leading to better health outcomes for clients

Thank you very much!

MACHINE LEARNING & ARTIFICIAL INTELLIGENCE

PROFESSOR: PABLO GERVAS GOMEZ-NAVARRO

Group 5:

Federico Barca, Felipe Fischel, Luca Conti, Ayça Basaran, Pablo Ostos