

# Recipe Site Traffic



Fernando Luque de la Torre

# Overview

**Goal:** Develop a model capable of predicting whether a recipe will lead to an increase in website traffic.

**Requirements set:** Be able to predict 80% of high traffic recipes.

**Tools:** Recipe data extracted by Product Manager (not all recipes data available)

# Work done

1. Data Validation:
  - a. Analyse different features
  - b. Dealing with missing values
2. Exploratory analysis
  - a. Check independence between features
  - b. Analyse distributions
  - c. Detect outliers
  - d. Analyse target variable and its dependence on different features
3. Development of the classification model
  - a. “Dummy” model
  - b. More complex model
  - c. Comparison with an even more complex model
4. Evaluation of the model

## Metrics obtained

- **Accuracy:** It shows how often a classification model is correct **overall**
- **Precision:** It shows how accurate the model is out of those predicted positive, how many of them are actual positive
- **Recall:** It shows how many of the actual positives the model captures

The metric requested by the Product Manager was the **recall** to be >80%

# Results obtained

## Dummy Classifier

Only remembers the majority class in each category.

- Test Recall: 67,12%
- Test Accuracy: 75,65%
- Test Precision: 73,13%

## Unbalanced Logistic Classifier

Gives more weight to the positive class to the detriment of the negative

- Test Recall: 82,19%
- Test Accuracy: 72,62%
- Test Precision: 62,5%

# Business Metrics

**What does it mean that we have managed to increase recall to 82%?**

It means that, of the total number of recipes that would drive traffic to the website, our model correctly predicts 82% of them.

**Now, what does it mean that the precision is 62.5%?**

It means that, of all the films that our model classifies as candidates to increase traffic to our website, only 62.5% of them will actually do so

The problem using the model is that it will show a lot of recipes that don't lead to high traffic.

# Final Summary and Considerations

Unsatisfactory results, perhaps due to the following factors:

- 1. Insufficient or Uninteresting Features**
- 2. Limited Data Quantity**

Possible future works:

1. Use more features such as the number of ingredients, the cost of the recipe or preparation time
2. Collect more data