Electric Cars

Machine Learning - EMlyon - 2021

Agenda

- Research definition
- Data collection & Process
- Results
- Key Learnings

Electric car : yes or not?

We have seen a trend toward electric cars, and we desire to launch our new product.

In order to predict the demand, we train diverse Machine Learning models based on Classification to understand If people would like to buy our new electric car and if yes what would be the ‘main important features. In the end, we chose the most accurate attempted to find a model that minimizes erros.

We are going to present you a summary of our data collection process, model implementation, and the best machine learning model adapted to our problem.

The process we followed...

CREATION OF THE SURVEY

DATA COLLECTION

DATA CLEANING

MODEL BUILDING AND TRAINING

CHOOSE THE BEST MODEL

The User Insights

Data Collection

1. SURVEY

We had a sample of 100 users in our target to be likely to purchase an electric car.

2. PREVIOUS FILTER

The survey was only shared with users holding a driving license because they are more likely to buy a new car.

3. INSIGHTS

16 Questions to obtain information with different variables regarding the main behavior of our target

User insights

Which are the main variables?

Categorical

- Main transportation
- Gender - (binary)
- Location - (binary)
- Energy

Numerical

- Quiet
- Speed
- Robustness
- Autonomy
- Design
- Technology
- Ecology

Numerical Interval

- Distance
- Frequency of usage
- Time having a car
- Age
- Income

Processing the data

UPDATE THE DATA

we changed some variable names.

TRANSFORMED INTERVALS

For some numerical variables (with intervals), we modified transformed the interval in its average

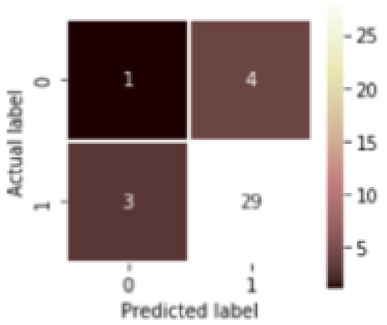
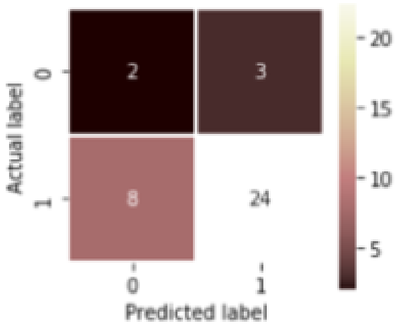
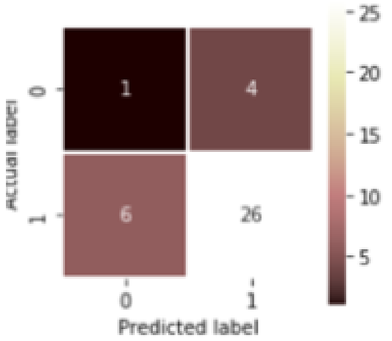
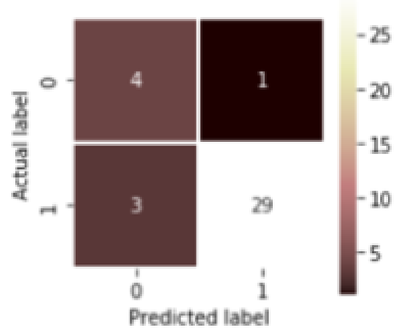
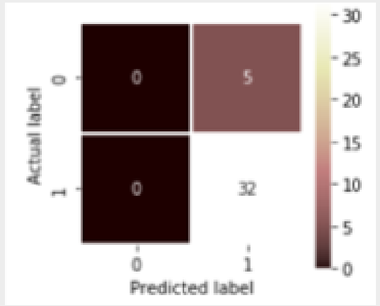
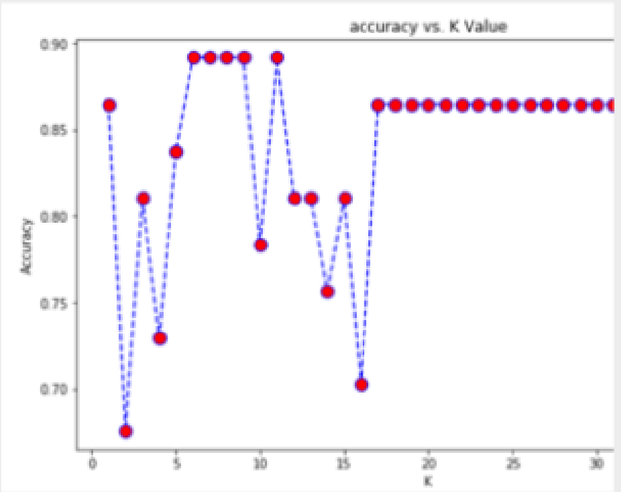
ENCODED VARIABLE

we encoded our categorical variable:
columns=*['Main_transportation', 'Energy', 'Gender', 'Location','Purchase_car']*

DIVIDED DATASET BETWEEN TRAINING SET AND TESTING SET

we chose a test_size of 0.35 and a random_state of 8 to get something more equilibrate (the most accurate)

Summary of our models

	SVC Kernel=linear	Decision tree	Logistic regression	KNearest Neighbours
Accuracy	81.1%	70.3%	73.0%	89.2%
Test matrices				
Comments	for a Kernel=poly The model has an accuracy of 86.5% But predicts that everyone will be buying electric cars. 	We changed the number of max iterations to get the best accuracy for this model. max_iter=10000	 n_neighbors=6 has the best accuracy	

Model choice

K-Nearest Neighbours

- Best accuracy
- 30 purchased predicted (only one is wrong)
- 7 non purchase predicted (4 are true)

- Based on training data: find the majority class of the k (here 6) closest neighboring data

ADVANTAGES

- Good method when accuracy is required (important for marketing investment)
- Good when you have data from different class (car purchase data is very diversified)
- Efficient
- Simple to understand

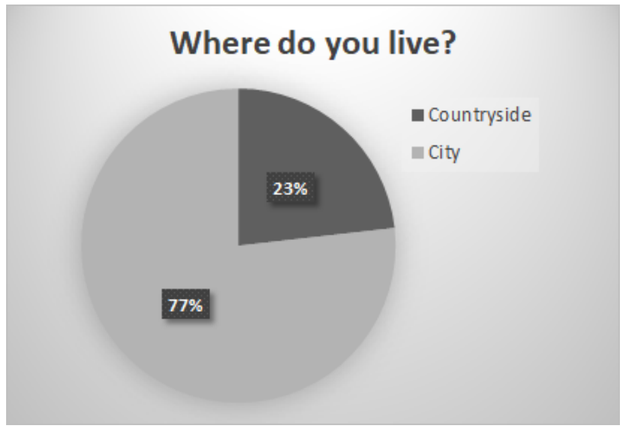
DISADVANTAGES

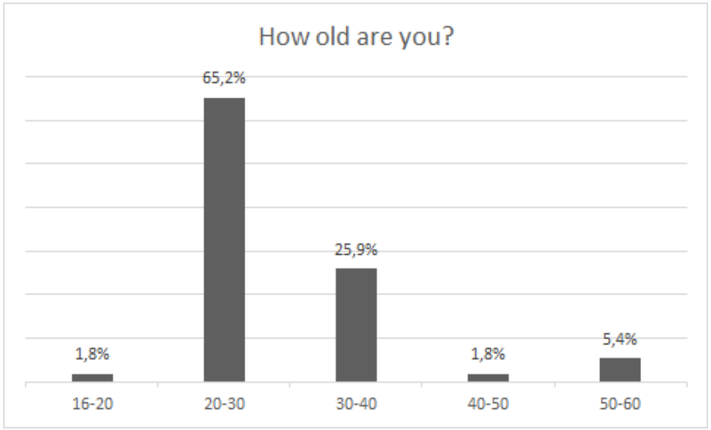
- Not good when it is supposed to be human readable
- Manual choice of the "k" number of neighbors
- Not a lot of data storage space

Next steps

LIMITS OF THE PROJECT

- Sample of only 105 observations
- Market representation is limited





REFINE AND REEVALUATE

- Add more data: more precision and more diversified observations (mass market representation)
- Keep verifying the accuracy of the model
- Maybe switch to another model (ex: decision tree more interpretable)

MARKETING CAMPAIGNS

- Use the data set to target people: Use our CRM data, past customer data, distribution partner data
- Promote the features that impact the more the choice of an electric car (make more modelling)

THANKS

BOSCQ MARIE
FONSECA DANIELA
MELLOULI MARIAM
PEREZ MUÑOZ LETICIA
SABBAH NAOMIE

EMLYON BS - MACHINE LEARNING 2021