Sales Forecasting for a pub - Telecom Bar'itech

Maxime MICHEL - Mathieu OLIVIER

https://github.com/Maxime-Michel-1999/BabarSalesForecast

1 Abstract

Based on a study made on Walmart Company we worked on predicting the sales of a pub. Our goal was to first find the features that influenced the consumption of beers in the pub and then to be able to predict the number of sales. By testing a lot of algorithms we found out that Random forest was the best way to predict. The result of our work is a table with the number of beer the user will sale for each week.

2 Introduction

For a business to be optimized it seems necessary to have an idea about how much is going to be sold, it enables the business house to produce or gather the right quantities at the right time. Furthermore it enables to make arrangement in advance for material, equipment or labor.
In a bar or restaurant it is very important to have an approximation of the futur sales as it allow the owner to order the right quantity of food and avoid any problem with waste or expiration (which can quickly become expensive for those kind of buisness). It is usefull both financially and ecologically. Any forecast can be termed as an indicator of what is likely to happen in a specified future time frame in a particular field. Therefore, the sales forecast indicates as to how much of a particular product is likely to be sold in a specified future period in a specified market at specified price. Our solution is to offer a trustful estimation of the future sales for a bar. This would accompany the manager in his/her choices about orders and organization. We would use multiple machine learning models to have the most precise forecast on a weekly scale, and this by only using a few information like the holiday or exam dates. Furthermore this forecast would help the management in determining as to how much revenue can be expected to be realized and what shall be the requirement of men, machine and money.
We worked on several ML models of regression with as input some characteristics of the period and as output the number of beer the user should buy for each category each week.

3 Related Work

Looking for examples on the web we've found an interesting work on the Walmart Company Sales *https://medium.com/analytics-vidhya/walmart-sales-forecasting-d6bd537e4904.*The data collected ranges from 2010 to 2012, from 45 Walmart stores across the country. Then, using this data, the goal was to predict the sale grouped by department. From this work we learned a lot about the ways to use our data and more on the methods and algorithms we may have to use.

4 Dataset

4.1 Get the Data

The first step was getting the Database from the pub. It's a collection of sales historic during 8 years. It was a MySQL database and it contained all the sales by drinks, customers and dates. Then we had to transform the data because there are too many products sold to have a prediction for each one. So we decided to gather them into categories. We want to gather them into pertinent groups so that the forecast can really help the bar.

4.2 Choosing the Products

The first step was to choose which product we were going to take account of. We chosed to focus our work on the beverages sales so we exclude any food or other kind of sales.Thanks to a SQL query and MariaDb we had a data table gathering only the sales for the products we had chosen to consider.

4.3 Grouping the items

The second step was to find groups to gather the beverage in. Indeed a majority of those products have been coming and leaving from the bar menu so their sales aren't usable one by one. Therefore we decided to group them into pertinent categories. We created the following 4 groups with MariaDb : High Degree Beer, Normal Degree Beer, Not Beer, Special Beer. The categories are based on the fact that every 1 or 2 week at least one product of each category was being sold, assuring the continuity of sales in each categories. The goal is having sales that are homogene on a year scale, meaning with no large unjustified void period.

4.4 Find a good time scale

We think that a Weekly prediction can be done, so we'll consider the sales week by week (Time series approach). We will also maybe consider a year index. Indeed promotion aren't all the sames therefore they don't consume the same quantities. Adding a bias for each promotion could maybe increase the accuracy of our estimations.

5 Features

In order to do our prediction we used the following features thinking that those are the main influences on whether people buy more or less drinks in a bar:

- Number of week : for each week we will sum the sales for each categories. The weeks will be registered by an index of week from 1 to 52. We used python to do that. The results is a dataframe with the sales for each category for each week.
- Promotion Index : As told before we decided to add a coefficient that reflects the promotion ! Every promotion doesn't drink the same, a coefficient should be schoolyear based. So we just add a feature indicating the "promotion".
- Week before next holiday : After thinking about it we decided to not only consider if the week is or isn't a holidays week but also count the number of week before the next holidays. Indeed we foresee that people tend to buy more in a Bar when a Holiday is near. So the data given to the machine learning model will be the counter to the next holidays.
- Week before next exams by a coefficient : As we did for the holidays what would interest us is a counter to the next exam. But in the case of exams we also want to transcript the possibility of several week of exams following each other. So our idea is to examine the 3 following weeks and ponderate them.

6 Methods

Now the core of our work is to choose the best method to make predictions from our data. Or goal is to look in the dataset for features such as trends, cyclical fluctuations, seasonality, and behavioral patterns. For this first version all the models where taken with their default parameters as there weren't some noticeable improvement when changing them. Here are the algorithm that we used to forecast sales :

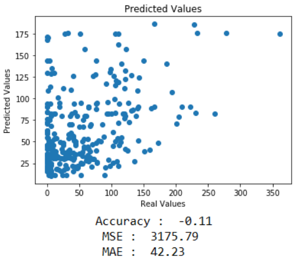
- KNN Regressor : this algorithm uses ‘feature similarity’ to predict the values of any new data points. This means that the new point is assigned a value based on how closely it resembles the points in the training set.
- Decision Tree : it identifies ways to split a data set based on different conditions.
- Random Forest : the Decision Tree algorithm has a major disadvantage in that it causes over-fitting. This problem can be limited by implementing the Random Forest Regression in place of the Decision Tree Regression. Additionally, the Random Forest algorithm is also very fast and robust than other regression models. To summarize in short, The Random Forest Algorithm merges the output of multiple Decision Trees to generate the final output.
- Extra tree regressor : this class implements a meta estimator that fits a number of randomized decision trees (a.k.a. extra-trees) on various sub-samples of the dataset and uses averaging to improve the predictive accuracy and control over-fitting.
- MLP regressor : this model is a neural net that optimizes the squared-loss using stochastic gradient descent.

7 Results and Discussion

7.1 Training and Test Data

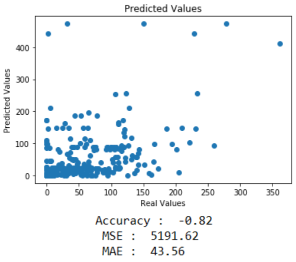
By trying those algorithm we found out that the way of choosing the training and test data had an import impact on the results. We first decided to go with the following repartition : 80% for the training set and 20% for the test. What was observed is that shuffling the data before splitting had a large impact on the results. Indeed taking the last year as testing data gives worse results than shuffled test data. It means that there are some factors that we should have add as features, that could be an improvement. There are the comparisons between shuffle and not for each model :

Not Shuffled



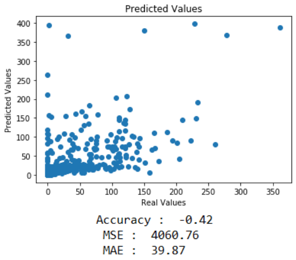
```
Accuracy :  -0.11
MSE :  3175.79
MAE :  42.23
```

Figure 1: KNN



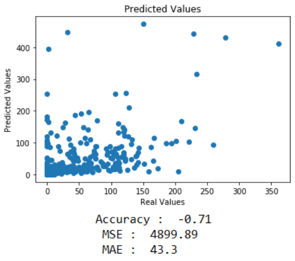
```
Accuracy :  -0.82
MSE :  5191.62
MAE :  43.56
```

Figure 2: Decision Tree



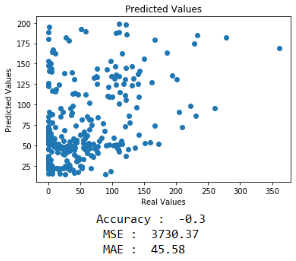
```
Accuracy :  -0.42
MSE :  4060.76
MAE :  39.87
```

Figure 3: Random Forest



```
Accuracy :  -0.71
MSE :  4899.89
MAE :  43.3
```

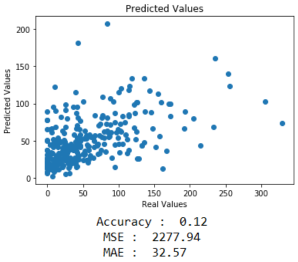
Figure 4: Extra Tree



```
Accuracy :  -0.3
MSE :  3730.37
MAE :  45.58
```

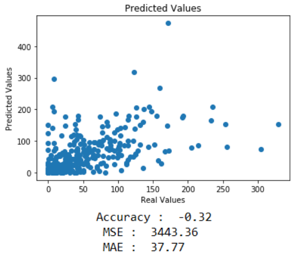
Figure 5: MLP regressor

Shuffled



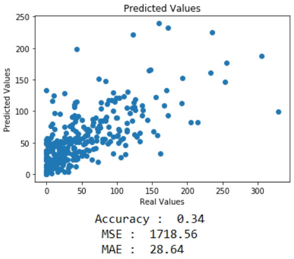
```
Accuracy :  0.12
MSE :  2277.94
MAE :  32.57
```

Figure 6: KNN



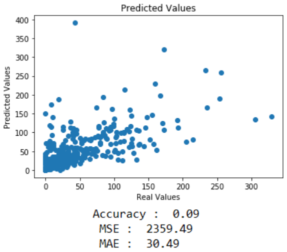
```
Accuracy :  -0.32
MSE :  3443.36
MAE :  37.77
```

Figure 7: Decision Tree



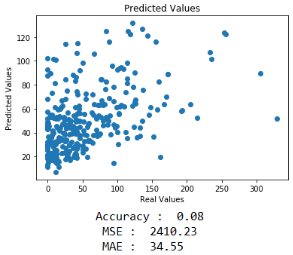
```
Accuracy :  0.34
MSE :  1718.56
MAE :  28.64
```

Figure 8: Random Forest



```
Accuracy :  0.09
MSE :  2359.49
MAE :  30.49
```

Figure 9: Extra Tree



```
Accuracy :  0.08
MSE :  2410.23
MAE :  34.55
```

Figure 10: MLP regressor

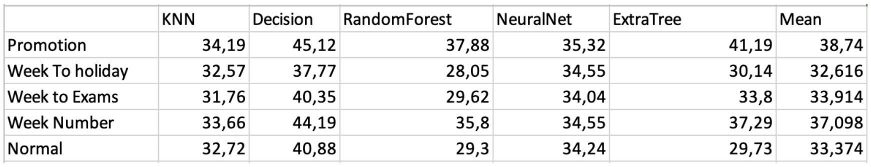
Interpretation The difference between shuffle and not shuffle is very notable ! So the models are missing something that considerate the year maybe ? We thought of adding the promotion as a feature would be a solution but apparently that's not enough. Shuffled data gives a better result but has no sense in term of prediction as the goal is for the owner to predict futur value.

7.2 The metric

In order to compare the models we need to focus on a metric. There are the three metrics that were considered : score (r^2) / Mean Absolute Error (MAE) / Mean Squared Error (MSE). We think that the best metric to use is the mean squared error (RMSE), but in a concrete way the best one is actually the absolute error which is the most indicative as it shows the actual difference that the bar will encounter from prediction to actual sales. We are working with a bar so we think that the best way to evaluate the best models is to consider the MAE only. However we understood that a real study on the metrics used could have a large impact on the work and improve the results !

7.3 Comparing Importance of features

In order to see the importance of features we decided to remove each one of them and observe the effect of thir removing. In the following table, the name of the feature corresponds to the one not used.



	KNN	Decision	RandomForest	NeuralNet	ExtraTree	Mean
Promotion	34,19	45,12	37,88	35,32	41,19	38,74
Week To holiday	32,57	37,77	28,05	34,55	30,14	32,616
Week to Exams	31,76	40,35	29,62	34,04	33,8	33,914
Week Number	33,66	44,19	35,8	34,55	37,29	37,098
Normal	32,72	40,88	29,3	34,24	29,73	33,374

Figure 11: Features MAE Comparison

Promotion : It seems to be a very important feature, when we take it out the results drops
Exams and holiday: It doesn't affect the results as much as promotion but it still taking them up, when dropping this feature we observe a small drop in the results.
Week Number: Very important, as promotion dropping it really worsen our results

It seems that some other features would be needed, like the schedule of events or the number of students of each year (we could have access to the number of clients and not only the sales).

7.4 Concrete Results

Using MAE as a metric we finally choosed the random forest model which was the most suited for our needs. When taking the last year as a test data (not shuffled) we obtained the following relsults. The blue line is the real data result. The red one is the prediction.

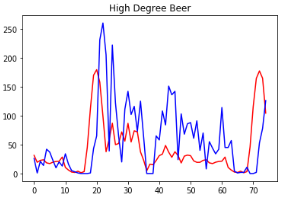


Figure 12: High Degree

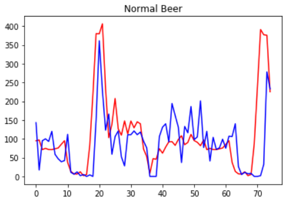


Figure 13: Normal Beer

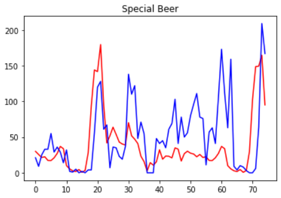


Figure 14: Special Beer

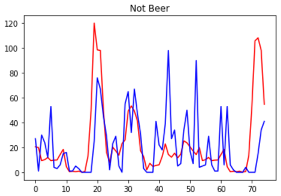
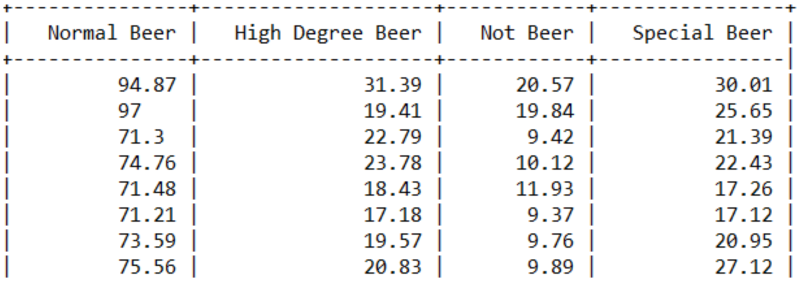


Figure 15: Not Beer

We can observed that the main patterns are recognized by our results. Nevertheless, many picks are not taking into count by our model which leads to a large amount of mistakes for several weeks. We see that the predictions are smoother than the real data, we think this is due to random events that our model doesn't consider at all. However we could still give advices to our user by giving him the fololowing table which is the final output of our work.



```
+---------------+-------------------+------------+---------------+
|   Normal Beer |  High Degree Beer |   Not Beer |  Special Beer |
+---------------+-------------------+------------+---------------|
|         94.87 |             31.39 |      20.57 |         30.01 |
|         97    |             19.41 |      19.84 |         25.65 |
|         71.3  |             22.79 |       9.42 |         21.39 |
|         74.76 |             23.78 |      10.12 |         22.43 |
|         71.48 |             18.43 |      11.93 |         17.26 |
|         71.21 |             17.18 |       9.37 |         17.12 |
|         73.59 |             19.57 |       9.76 |         20.95 |
|         75.56 |             20.83 |       9.89 |         27.12 |
```

Figure 16: Output expected for a user

This table is very easy to use, the numbers conrresponds with the quantities of beer for each category for each week. We would also notice the user that the MAE is of about 28 sales per week. It means that the gap between the predicted value and the reality could be of about 28 sales.

8 Conclusion

Seeing the graphs shows that the considerable irregularities in the sales of the establishment creates real difficulties when trying to forecast. Indeed our models are not able to forecast the drops or pics in sales that are due to some random particular event that takes place often in an university environment. However the random forest model still propose some notable results with a MAE of 28, with means that in average the model differs of 28 sales each week, which isn't that bad ! Knowing that those product have long expired dates then it isn't a real problem to have +-28 beers at the end of the week ! We could even see if it equalize over time (maybe the week after it compensates).