Contents

1 Introduction

Within the subject of Artificial Intelligence Techniques in Forecasting and Optimization in Business Systems, we were proposed a project aimed at using forecasting and optimization techniques in a real mind problem, in this case the distribution of drinks in a company.

Throughout this document we will first analyze the beverage sales data, then we will apply forecasting techniques to predict the number of sales of each beverage, this forecast will be divided by a univariate forecast and a multivariate forecast, for both will be used the knowledge and practices acquired in class to predict successfully. Finally, we have Optimization, which consists of finding the best values in order to maximize the sales value of each drink.

2 Project Execution

This section of the document is to describe the work done by each of the elements in the group. Each description will include each of the tasks completed by each group, the effort done for this work and the time spent doing this project.

For this work, we used a work methodology where, during the various meetings we held, we usually divided them into 3 parts. The first part consisted of analyzing the results of each of the team members, in case there were tasks to be done for that meeting. The second part of the meeting consists of analyzing what needs to be done in general for the next delivery, in that same time we visualized and aligned all the tasks to be done for the next presentation. Finally, the third and final part of the meeting consisted of dividing these tasks previously defined by each of the elements of the group.

We usually have at least 2 meetings in the space from one class to another, but there were situations where more than one meeting was necessary in that space of time, mainly due to doubts in the execution of the tasks, in these situations we normally gathered everyone by video call and resolved let's get to the problem.

2.1 Work Done By Dong Xuyong

For this project the tasks made by this member where:

•Extract business objectives into features;

•Model prediction with LSTM;

•Reading the article “Understanding LSTM”;

•Run all rminer ML models;

•Parameter tuning for bud and stella;

•Pipeline for all model types;

•Analyze and run all models with all metrics for univariate variables, with different timelag combinations;

•Growing and Rowling window;

•Multivariate with VAR model and Arima with exogenous variables (precipitation and temperature);

•Weakly Naive probability implementation and experience;

•Fix the Weekly Naive template;

•Implement the GW with neural networks with multivariate series and 2 outputs and model tuning in Python;

•Analyze the optimization method.

This member of the group spent around 68 hours in this project;

2.2 Work Done By Pedro Silva

For this project the tasks made by this member where:

•Extract business objectives into features;

•Analysis of the exponential smoothing algorithm;

•Research in GW and RW methods;

•Search R tools;

•Formulate the validation function;

•See validation algorithms;

•Think of a strategy on how to implement this same;

•Implementation of the optimization method;

•Interface implementation.

This member of the group spent around 65 hours in this project;

2.3 Work Done By Tiago Martins

For this project the tasks made by this member where:

•Extract business objectives into features;

•Analysis of Forecasting with Holt-Winters;

•Research in GW and RW methods;

•Final document development;

•See validation algorithms;

•Implementation of the optimization method;

•Interface implementation.

This member of the group spent around 60 hours in this project;

2.4 The Work Methodology and Auto-Evaluation

For the Group Auto-Evaluation we think that we deserve **16** for our project final grade. We think that we deserved this grade because we where able to complete the tasks:

•Dataset description;
•Univariate prediction using train-test split;
•Univariate prediction using Growing and Rowling window split;
•Multivariate prediction;
•Otimization of transport;
•Interface.

And there where no tasks that we could not completed, although we believe that we could do a better job with the interface.

For the Individual Auto-Evaluation our group thinks this the grade that each element of the group deserves:

DONG	PEDRO	TIAGO
16	16	16

Figure 1: Group Auto-Evaluation

3 Dataset Description

For this project, we were provided with an excel file called "bebidas.xlsx", within which are the daily sales records of each of the two beverages made available by the company in question, within that excel file there are still other relevant data, which will be detailed afterwards.

In the following image we have a print of the columns of the dataset mentioned above:

	A	B	C	D	E	F	G
1	DATA	DIA_SEMAN/	PRECIPITACA	TEMP_MAX	STELLA	BUD	
2	01/02/2019	4	6,8	30,1	53	71	
3	01/03/2019	5	0	32,9	106	235	
4	01/04/2019	6	14,2	31,8	218	42	
5	01/05/2019	7	3	27,7	180	110	
6	01/06/2019	1	0,6	29	69	15	
7	01/07/2019	2	0	31,6	18	8	
8	01/08/2019	3	0	33,2	61	10	
9	01/09/2019	4	0	31,1	38	6	
10	01/10/2019	5	0	33,2	545	26	

Figure 2: Project Dataset

The following dataset is compose of 6 columns, they being:

•DATA: This column represents the date the records are from;

•DIA_SEMANA: This column represents the day of the week, where 1 is Sunday, 2 is Monday, 3 is Tuesday, 4 is Wednesday, 5 is Thursday, 6 is Friday and 7 is Saturday;

•PRECIPITACAO: This column represents the total of precipitation in mm in that day;

•TEMP_MAX: This column represents the daily maximum temperature in Celcius from that day;

•STELLA: This column represents the number of STELLA drinks that where sold in that day;

•BUD: This column represents the number of BUD drinks that where sold in that day.

To get a nice description of the values of each column we calculate the minimum value, the median, the mean, the max and other important values for each column, the next image represents the values that we got:

DATA	DIA_SEMANA	PRECIPITACAO	TEMP_MAX	STELLA	BUD
Min. :2019-01-02	Min. :1.000	Min. : 0.000	Min. :21.40	Min. : 0.0	Min. : 0.0
1st Qu.:2019-07-03	1st Qu.:2.000	1st Qu.: 0.000	1st Qu.:29.20	1st Qu.: 13.0	1st Qu.: 22.0
Median :2020-01-01	Median :4.000	Median : 0.000	Median :30.90	Median : 47.0	Median : 58.0
Mean :2020-01-01	Mean :4.001	Mean : 3.669	Mean :31.13	Mean : 105.4	Mean : 101.4
3rd Qu.:2020-07-01	3rd Qu.:6.000	3rd Qu.: 1.550	3rd Qu.:32.80	3rd Qu.: 128.8	3rd Qu.: 125.8
Max. :2020-12-31	Max. :7.000	Max. :66.000	Max. :40.50	Max. :1335.0	Max. :1280.0

Figure 3: Dataset Statistics

These two diagrams represent the sales of each of the drinks provided in the dataset:

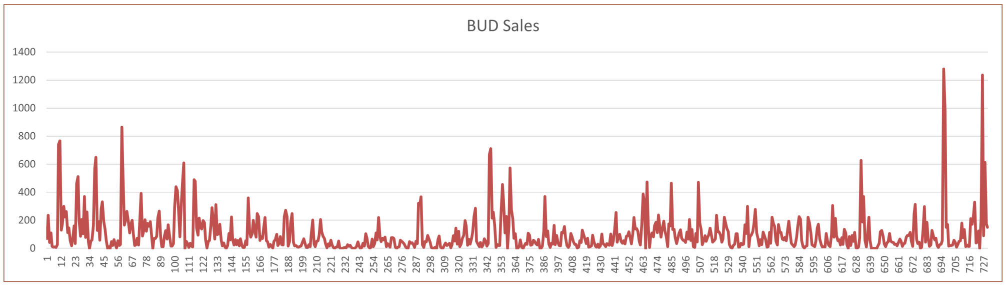

Figure 4: BUD Sales

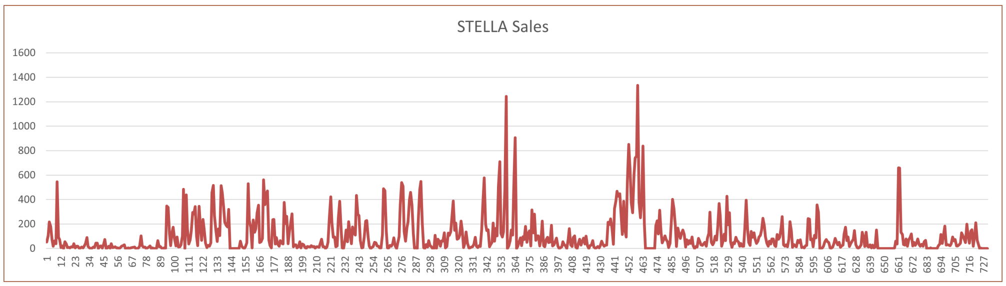

Figure 5: STELLA Sales

The next diagrams represent the outliers of sales data in realtion to the average sales of the respective dring (BUD and STELLA):

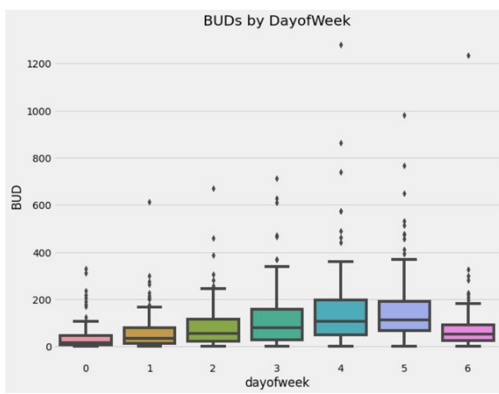

Figure 6: BUD Outliers

In this image we can see that int the BUD sales, there are more outliers in the Sunday and in Friday, but there are also some smaller outliers in the rest of the week, excluding Monday.

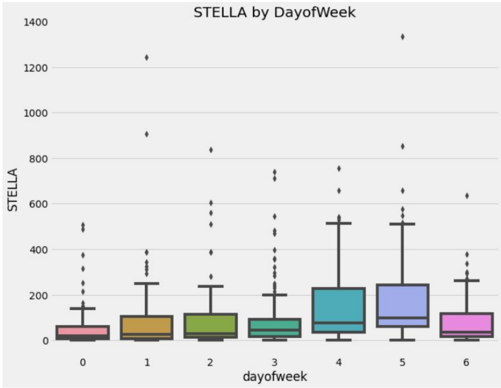

Figure 7: STELLA Outliers

In this image we can see that int the STELLA sales, there are more outliers in the Saturday and in Tuesday, but there are also some smaller outliers in the rest of the week, excluding Monday and Sunday.

The next two images will show the ACF of the STELLA sales and the BUD sales:

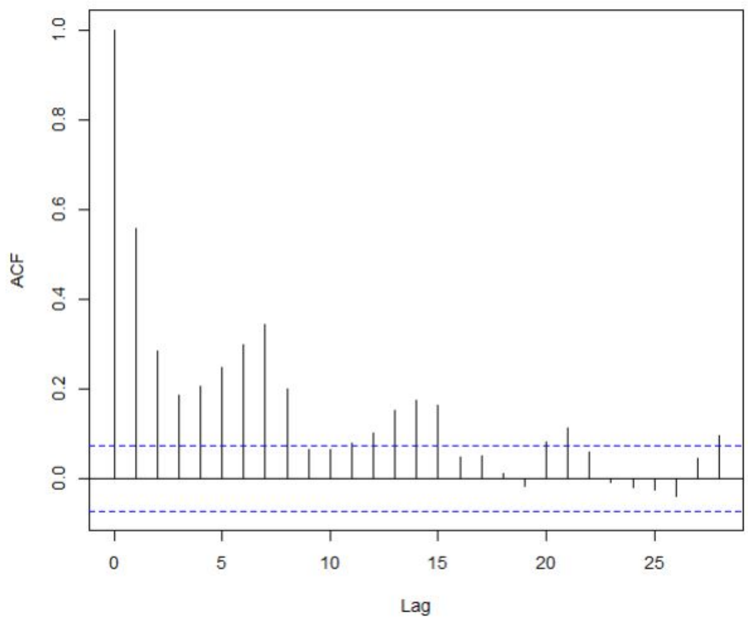

Figure 8: ACF of STELLA

For the ACF of the STELLA drink we can see that there is a pattern, the beginning of the data shows a peak in the ACF value, and it goes down for the next 4 results and then it increases for the next 4 values, and after it goes down for the next 3 values, this pattern repeates for the rest of the data to the extent that the value of the ACF reaches negative values.

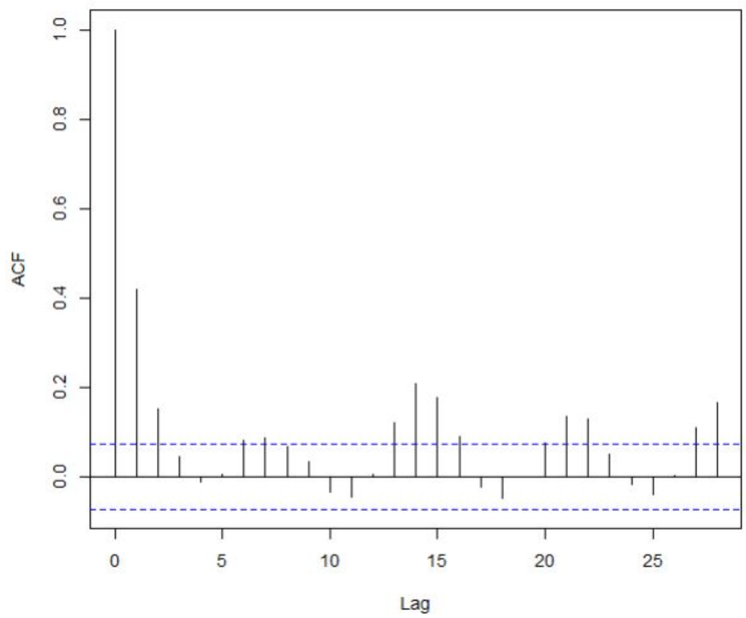

Figure 9: ACF of BUD

For the ACF of the BUD drink we can see that there is a pattern, the beginning of the data shows a peak in the ACF value, and it goes down for the next 5, reaching negative value of ACF in the last result, and then it increases for the next 3 values, and after it goes down for the next 4 values, reaching again negative values, this pattern repeates for the rest of the data.

The next image will show the correlation of each one of the drinks with the rest of the columns values:

```
main_df.corr()["BUD"].sort_values()
```

```
TEMP_MAX       -0.047237
PRECIPITACAO    0.004209
STELLA          0.092010
DIA_SEMANA      0.278348
BUD             1.000000
Name: BUD, dtype: float64
```

+ Code + Markdown

```
main_df.corr()["STELLA"].sort_values()
```

```
PRECIPITACAO   -0.033880
TEMP_MAX        0.005533
BUD             0.092010
DIA_SEMANA      0.234104
STELLA          1.000000
```

Figure 10: Correlation of STELLA and BUD

We can see from the correlation that for the two drinks they have some characteristics similiar, for example it shows that if the maximum temperature increases that means that there will less sales of both drinks, that because the value of the correlation is negative. In terms of precipitation we can see that when it rains more the value of the sales is biggers, for both drinks.

4 Prediction Objective

One of the tasks for this project is the prediction of the number of sales of each one of the drinks (STELLA and BUD). To do that task there are two main modules that we can use, Univariate Analysis and Multivariate Analysis.

Univariate Analysis consiste in examening the relationship between a single column of data, that means that for this project we will create 2 univariate predictions, one for STELLA and the other for BUD. Also each one of those predictions will include multiple prediction methods, and the objective is to find the predictionmethod with the lowest error rate, for each of the drinks.

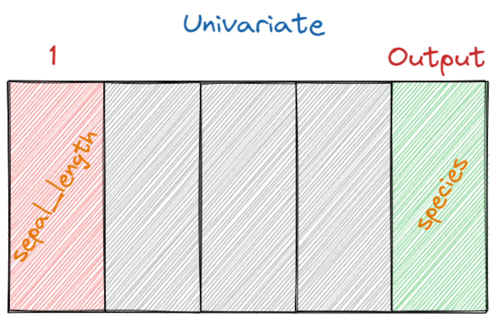

Figure 11: Univariate Analysis

Multivariate Analysis is similar to the univariate analysis the main difference between the two is that multivariate analysis does not focus in only onde column of data, but instead use the data from multiple columns of data to predict. The data that will be included for the prediction of the sales of eac hdrink will include of course the sales of each respective drink, the precipitation and the maximum temperature on that day.

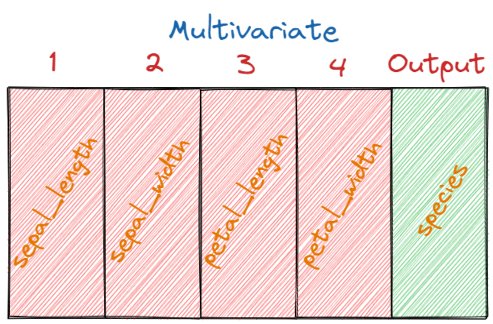

Figure 12: Multivariate Analysis

Across all the predictions we will be using a seed of 24 to ensure consistent and reproducible results in the machine learning models.

4.1 Univariate Analysis

For the Univariate Analysis there are multiple prediction methods that we will use, we will use mainly the machine learning amd forecast methods.

Our objective was to predict the last 20 weeks of each one of the drinks, for that we will use two methods for training the machine, they being the train-test split and the Growing and Rowling window split.

4.1.1 Train-Test Split

In the Train-test split as we metion above, we will use Machine Learning prediction methods and Forecast prediction methods to predict the last 20 weeks of the sales. For each methods we create a script that run all of them and later will show the best methods. To determine the best method we will calculate the average MSE error value from each method and of course the method with the lowest average MSE error will be considerd the best prediction method.

For the Machine Learning prediction methods we will use the following:

•"naive";

•"ctree";

•"cv.glmnet";

•"kknn";

•"mlp";

•"randomForest";

•"xgboost";

•"cubist";

•"lm";

•"mars";

•"pcr";

•"plsr";

•"cppls";

•"rvm".

For the Forecast predictions methods we will sue the following:

•"Holt-Winters";

•"auto.arima";

•"ets";

•"nnetar".

For the split we use the follow configurations:

•The size of the data in question is 730;

•After the time series transformation the size becomes 723;

•The trainning set consist of 85% of train set and the rest 15% is the validation set;

•The test set consists of 20 runs with the step of 7;

•The size of the window is 20.

After running the script that was developt by our group, we determine that the best prediction method was Holt Winters for the BUD drink, and mars for the STELLA drink, as they whete the methods with the lowest error rate from all the Machine Learning predictions methods and all the Forecast predictions methods. The next imagem will allows us to see the all the error rates from all the predictions methods.

	Model	RMSE
19	HW	91.2
20	auto.arima	94.5
22	nnetar	96.2
21	ets	97.1
8	mlpe	97.8
9	randomForest	99.5
6	ksvm	106.0
10	xgboost	107.9
11	cubist	109.2
12	lm	109.2
13	mr	109.2
15	pcr	109.2
16	plsr	109.2
17	cppls	109.2
1	naive	109.8
3	cv.glmnet	110.1
4	rpart	124.9
18	rvm	125.6
2	ctree	141.3
14	mars	142.5
7	mlp	163.9
5	kknn	198.2

Figure 13: BUD - Machine Learning and Forecast Results

As we can see from this image, the best prediction method for the BUD drinks was Holt Winters, also the second best and the thir best methods were auto.arima and nmetar. Just a side note we can see that the best 3 methods come from Forecast predictions.

	Model	RMSE
14	mars	185.8
1	naive	186.5
12	lm	186.5
13	mr	186.5
15	pcr	186.5
16	plsr	186.5
17	cppls	186.5
3	cv.glmnet	186.6
9	randomForest	186.7
2	ctree	187.3
4	rpart	187.4
7	mlp	187.9
8	mlpe	189.2
18	rvm	190.7
21	ets	190.8
20	auto.arima	194.4
11	cubist	198.2
10	xgboost	201.0
5	kknn	201.3
6	ksvm	204.3
22	nnetar	206.1
19	HW	206.7

Figure 14: STELLA - Machine Learning and Forecast Results

As we can see from this image, the best prediction method for the STELLA drinks was mars, also the second best and the third best were naive and lm. Just a side note we can see that the best3 methods where form Machine Learning predictions.

For demonstration purposes here are the graphics of each of the predictions made by the best method for BUD and STELLA:

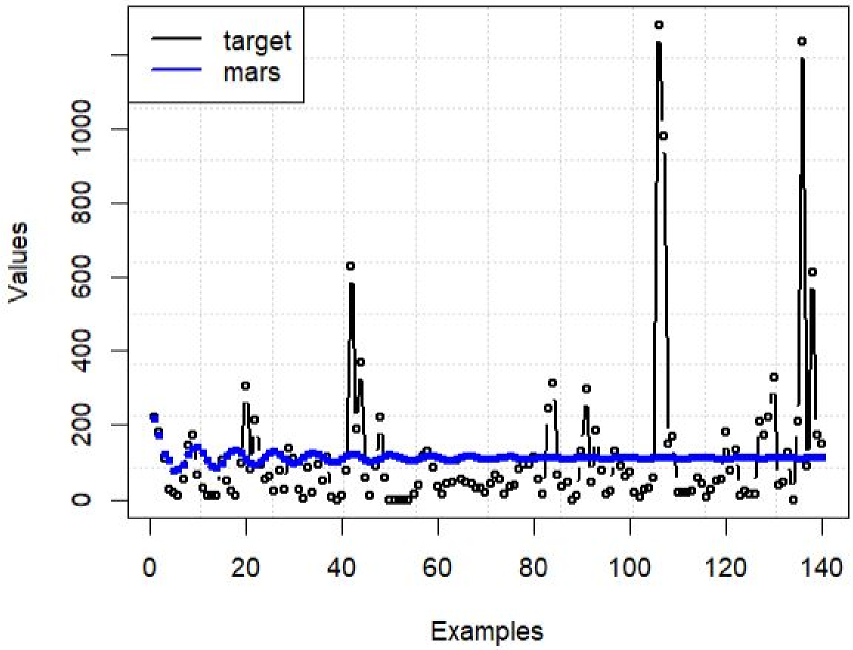

Figure 15: BUD - Machine Learning and Forecast HW Graph

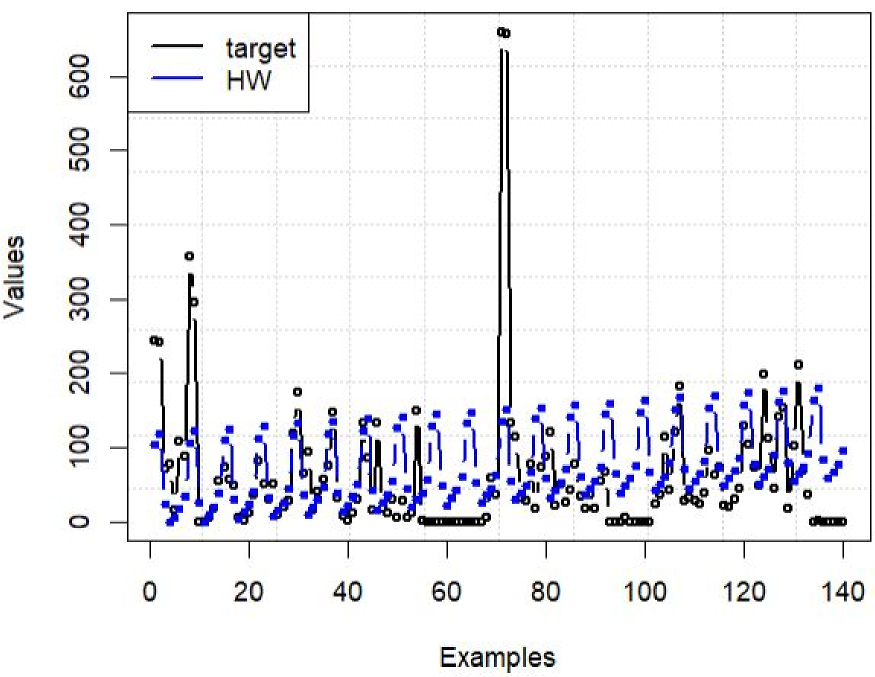

Figure 16: STELLA - Machine Learning and Forecast mars Graph

4.1.2 Sliding Window

To choose which split was the most efficient we calculate between the Growing window and the Rowling window which one we would choose:

Method	Mean of MSE (BUD)	Mean of MSE (STELLA)
GW	**21093.1**	**5194.4**
RW	28968.1	20195.8

Figure 17: Sliding Window Evaluation

In this image we test each of the two split methods with the ML Model, and we concluded that the Growing window split shows the best results.

Timeseries	Mean of MSE (BUD)	Mean of MSE (STELLA)
1:7	21093.1	**5194.4**
1:14	**13861.5**	8646.8
1,2,3,7	21738.0	5691.3

Figure 18: Time Series Comparison

In this image we compare the different time series with the ML model and we came to the conclusion that the time series 1:7 is best for STELLA and the time series 1:14 is best for BUD.

Similiar to above for this split we will use the same predictions methods, but we will also add the Weekly naive method.

```
> bud_metrics
```

	X	Model	MSE
1	16	lm	21093.14
2	17	mr	21093.14
3	19	pcr	21093.14
4	20	plsr	21093.14
5	21	cppls	21093.14
6	10	ksvm	21807.55
7	13	randomForest	22132.54
8	15	cubist	22925.25
9	18	mars	23605.71
10	5	naive	23827.39
11	7	cv.glmnet	24078.55
12	14	xgboost	24822.12
13	6	ctree	25071.61
14	12	mlpe	25989.65
15	9	kknn	25993.33
16	11	mlp	26054.86
17	8	rpart	27084.84
18	22	rvm	28047.55
19	3	ets	36382.69
20	2	auto.arima	37431.72
21	4	nnetar	40482.54
22	1	HW	41338.73
23	23	weekly_naive	46756.09

Figure 19: BUD - Growing Window Results

As we can see from this image, the best prediction method for the BUD drinks was lm, also the second best and the thir best methods were mr and pcr.

```
> stella_metrics
    X        Model       MSE
1  16           lm  5194.414
2  17           mr  5194.414
3  19          pcr  5194.414
4  20         plsr  5194.414
5  21        cppls  5194.414
6  18         mars  5880.614
7  12         mlpe  6323.571
8  15       cubist  6458.479
9  10         ksvm  6466.993
10 11          mlp  7026.807
11 14      xgboost  8163.121
12 13 randomForest  8236.414
13  2   auto.arima  8652.407
14  8        rpart  9006.429
15  6        ctree  9236.121
16  1           HW  9522.021
17 22          rvm  9697.829
18  9         kknn 10160.043
19  3          ets 10823.307
20  7    cv.glmnet 10852.136
21  5        naive 11267.979
22 23 weekly_naive 16219.343
23  4       nnetar 20830.221
```

Figure 20: STELLA - Growing Window Results

As we can see from this image, the best prediction method for the STELLA drinks was lm, also the second best and the third best were mr and pcr.

For demonstration purposes here are the graphics of each of the predictions made by the best method for BUD and STELLA:

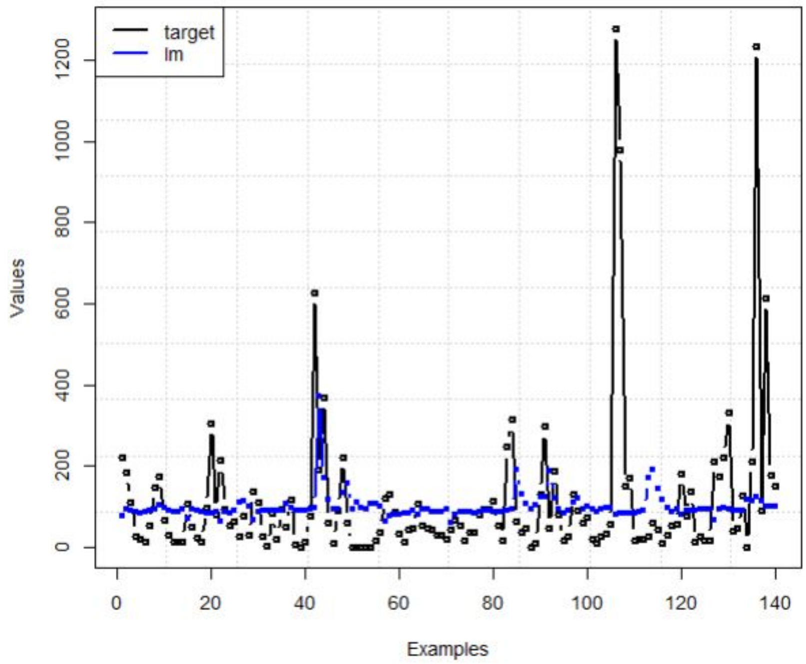

Figure 21: BUD - Growing and Rowling Window lm Graph

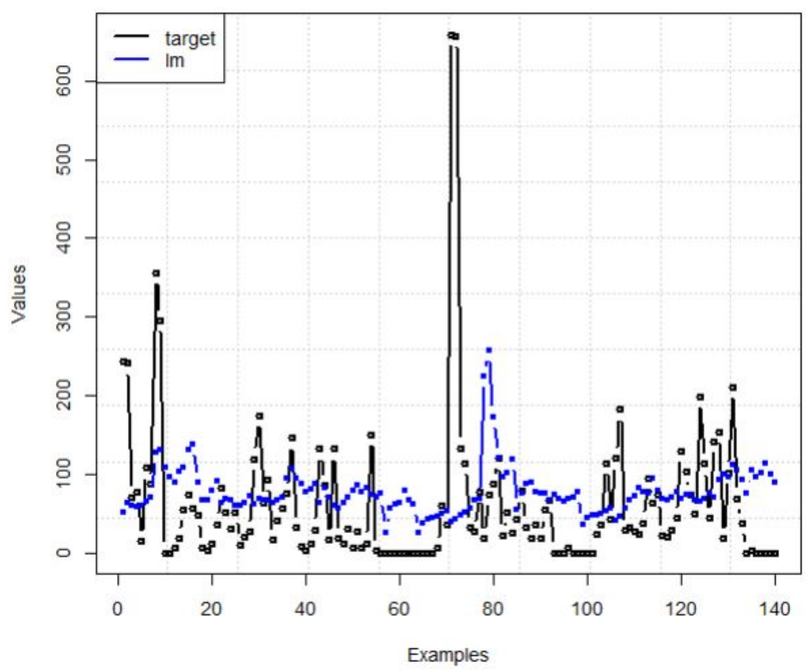

Figure 22: STELLA - Growing and Rowling Window lm Graph

4.2 Multivariate Analysis

The Multivariate Analysis is similiar to the Univariate Analysis, the diference is that in this one we will use more than a set of data, in this project we will use the precipitation and the maximum temperature to assiste in the prediction of sales for each of the drinks.

We try multiple neural networks architectures, with LSTM, to find which is the most efficient for our project:

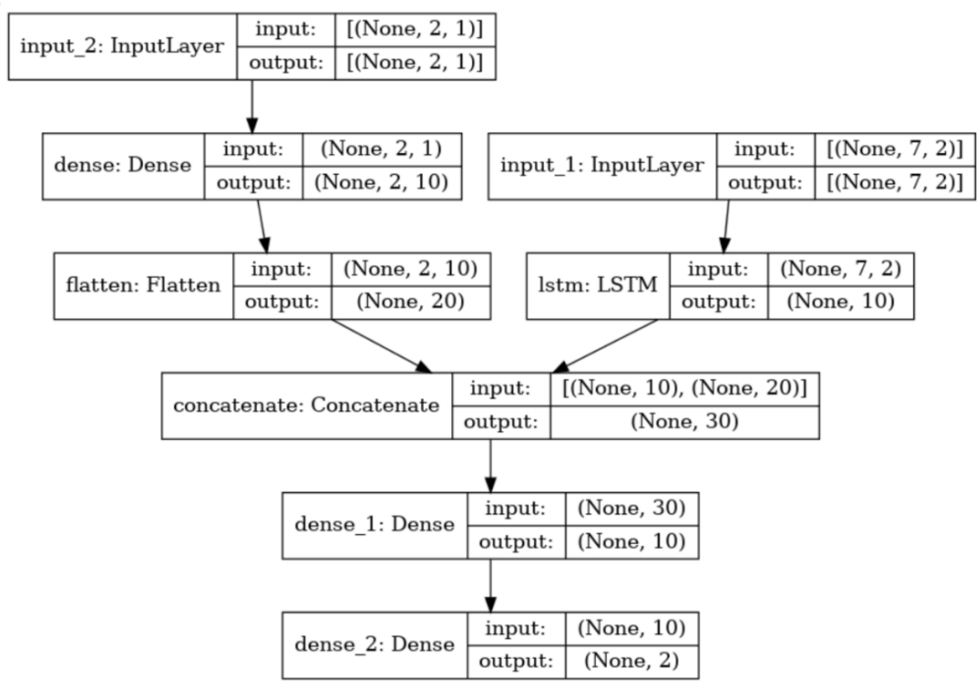

Figure 23: Neural Network architecture 1

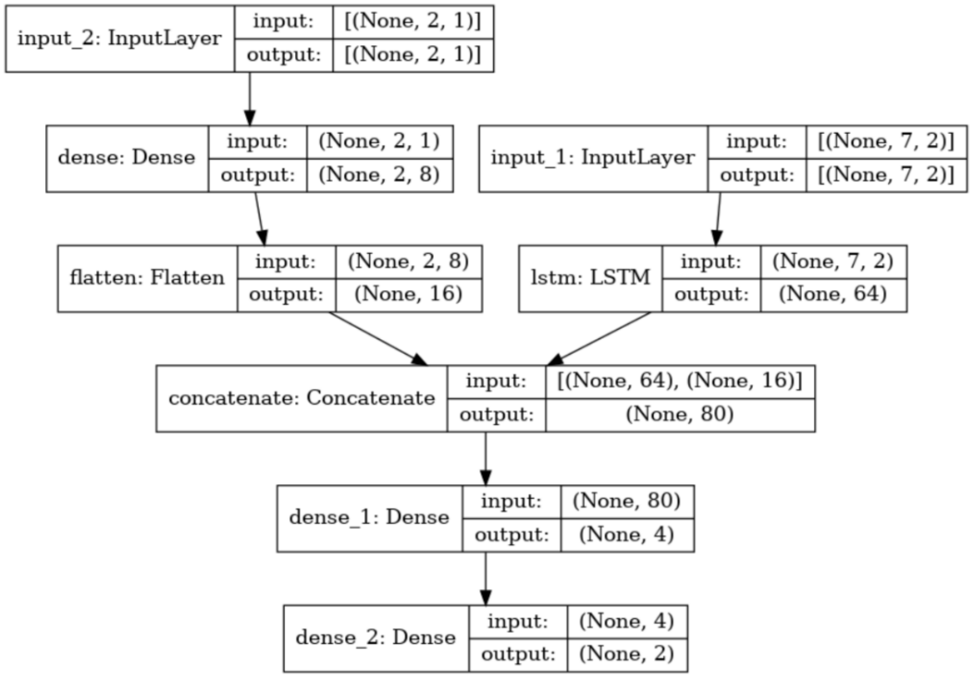

Figure 24: Neural Network architecture 2

Figure	Mean of MSE (BUD)	Mean of MSE (STELLA)
1	8095.0	6504.4
2	**6283.4**	**3542.6**

Figure 25: Neural Network Average MSE

As we can see from the image above, we conclude that the best neural network architectures is the second one.

We tried another approach with the following architecture:

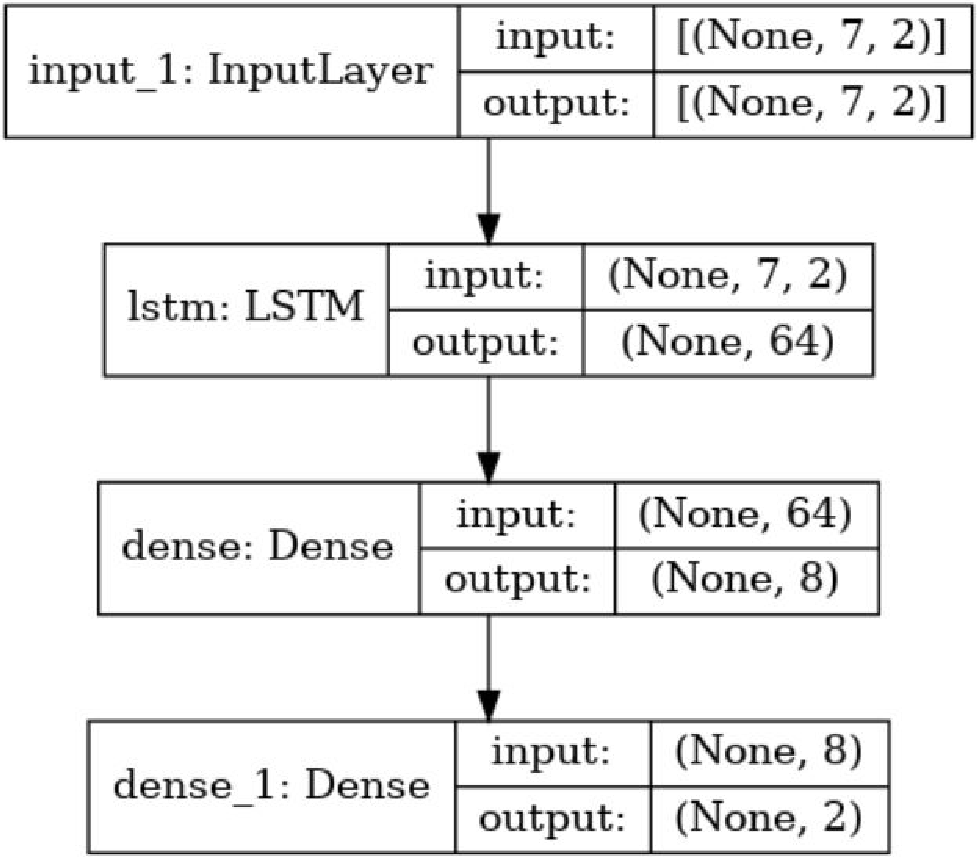

Figure 26: Alternative Neural Network

Learning Rate	Mean of MSE (BUD)	Mean of MSE (STELLA)
0.1	6510.2	7099.9

Figure 27: Alternative Neural Network Average MSE

As we can see the approach before have better results than the alternative.

For the prediction we will use a split of Growing Window to predict the last 20 weeks of sales, for each drink, with a step of 7. The next image represents the results of the average error MSE of the 20 weeks prediction:

Learning Rate	Mean of MSE (BUD)	Mean of MSE (STELLA)
0.001	5701.0	5930.5
0.03	**6283.4**	**3542.6**
0.01	6400.9	6566.0
0.1	6374.2	4788.0
0.3	7604.4	6282.2

Figure 28: Average of MSE Error for STELLA and BUD

For demonstration purposes here are the graphics of each of the predictions made for the BUD and STELLA:

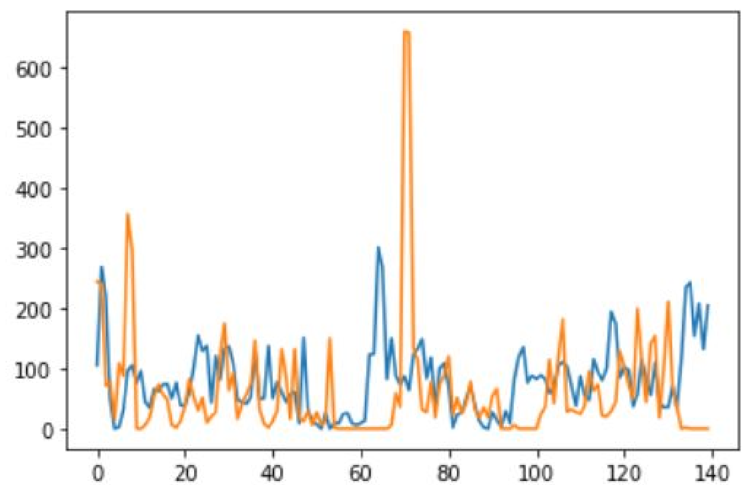

Figure 29: STELLA - Multivariate Analysis Results

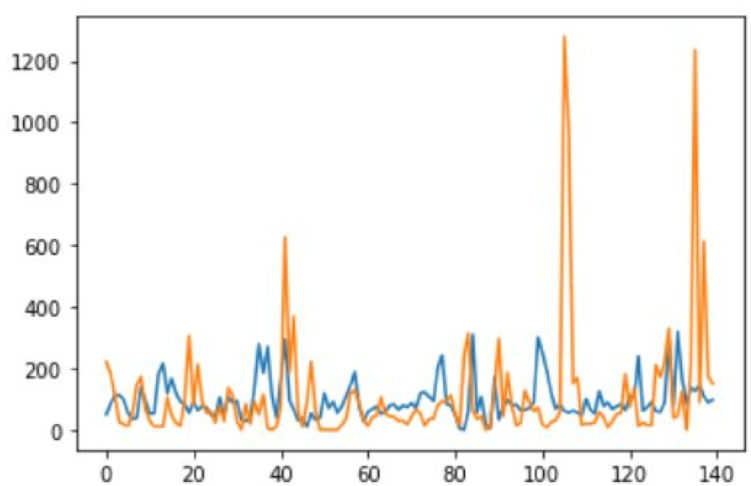

Figure 30: BUD - Multivariate Analysis Results

For last step we use the Weekly Naive as the standard metric to compare wwith the best result:

Model	Mean of MSE (BUD)	Mean of MSE (STELLA)
NN (lstm)	6282.4	3542.6
Weekly Naive	46756.1	26219.4
Performance	744%	740%

Figure 31: Weekly Naive Results

As we can see the STELLA results are 7.4 times better and the BUD results are 7.44 times better.

5 Otimization Objective

In the process of optimization, we undertook the following steps to improve our approach:

Initially, we developed an evaluation function based on the statement provided by the professor. This function served as the foundation for our optimization efforts. However, we soon realized that our initial implementation was not aligned with sound software engineering practices.

To address this, we decided to refactor the function. Our primary objective was to introduce flexibility by allowing the specification of the number of days to evaluate. Unfortunately, due to the nature of the input, we were always limited to 7 days.

What set our implementation apart from others was our approach to reducing the number of variables. We concluded that since most algorithms generated values randomly, we could streamline the process by utilizing only six variables. These variables, namely bud v1, bud v2, bud v3, stella v1, stella v2, and stella v3, represented the quantities of beer per brand allocated to different resources $v1, v2, v3$. We found no constraints in the provided statement that prohibited placing different beer brands within the same resource.

By leveraging this reduced set of variables, we were able to infer the remaining values based on the output generated by the respective algorithms. The analogy of playing dominoes accurately describes this approach, as each value falls into place based on the preceding one.

This methodology significantly minimized the need for value repairs. Since the values were inferred in accordance with the algorithm's output, we encountered fewer instances where, for instance, there were 100 beers in v1 resource but none in storage—a highly improbable scenario.

Furthermore, we established that a value range of 0 to 100 was suitable for the resources. Given that the total quantity of beers did not exceed 200, this range provided a reasonable and logical representation. Additionally, it allowed for sending a maximum of 300 units in a single day per resource, which surpassed practical requirements. By adhering to this range, the randomly generated values aligned with our expectations, ensuring meaningful results.

Initially, our focus revolved around maximizing profits, which required optimizing the entire function. To achieve this, we employed various algorithms, including hill climbing and Monte Carlo simulation.

As we shifted our attention towards minimizing the costs, we explored additional algorithms such as simulated annealing, SANN, grid search, and tabu search. However, after thorough evaluation, we concluded that Monte Carlo simulation and hill climbing emerged as the most effective algorithms for our specific problem. The alternative approaches failed to produce satisfactory results when evaluated against our optimization objectives.

This report extends the analysis of the Hill Climbing and Monte Carlo algorithms' performance for beer production optimization by comparing the optimal results achieved in Week 3 and Week 20. The focus is on identifying the best outcomes in terms of production predictions and evaluation scores for each algorithm in these two weeks.

Week 3 Analysis:

In Week 3, both the Hill Climbing and Monte Carlo algorithms were subjected to 2000 iterations. The optimal results obtained by each algorithm are as follows:

Hill Climbing:

- Stella Prevision: 119, 167, 119, 91, 85, 55;
- Bud Prevision: 62, 73, 74, 50, 76, 39;
- Evaluation Score: 3879.2.

Monte Carlo:

- Stella Prevision: 119, 167, 119, 91, 85, 55;
- Stella Prevision: 119, 167, 119, 91, 85, 55;
- Bud Prevision: 62, 73, 74, 50, 76, 39;
- Evaluation Score: 3676.1.

Comparing the algorithms, it is evident that both Hill Climbing and Monte Carlo produced identical predictions for Stella and Bud beer production in Week 3. However, the evaluation score favored the Monte Carlo algorithm, indicating its superior performance in optimizing the production levels during this week.

Figure 32: Multivariate Analysis

Figure 33: Multivariate Analysis

Week 20 Analysis:

In Week 20, once again using 2000 iterations, the algorithms' optimal results were as follows:

Hill Climbing:

•Stella Prevision: 84, 137, 127, 145, 109, 89;

•Bud Prevision: 115, 234, 243, 154, 207, 132;

•Evaluation Score: 8471.8.

Monte Carlo:

•Stella Prevision: 84, 137, 127, 145, 109, 89;

•Bud Prevision: 115, 234, 243, 154, 207, 132;

•Evaluation Score: 8518.1.

Analyzing the optimal results in Week 20, both algorithms once again produced similar predictions for Stella and Bud beer production. However, the evaluation scores indicate a marginal advantage for the Monte Carlo algorithm, suggesting its slightly better performance in optimizing production levels during this week.

Figure 34: Multivariate Analysis

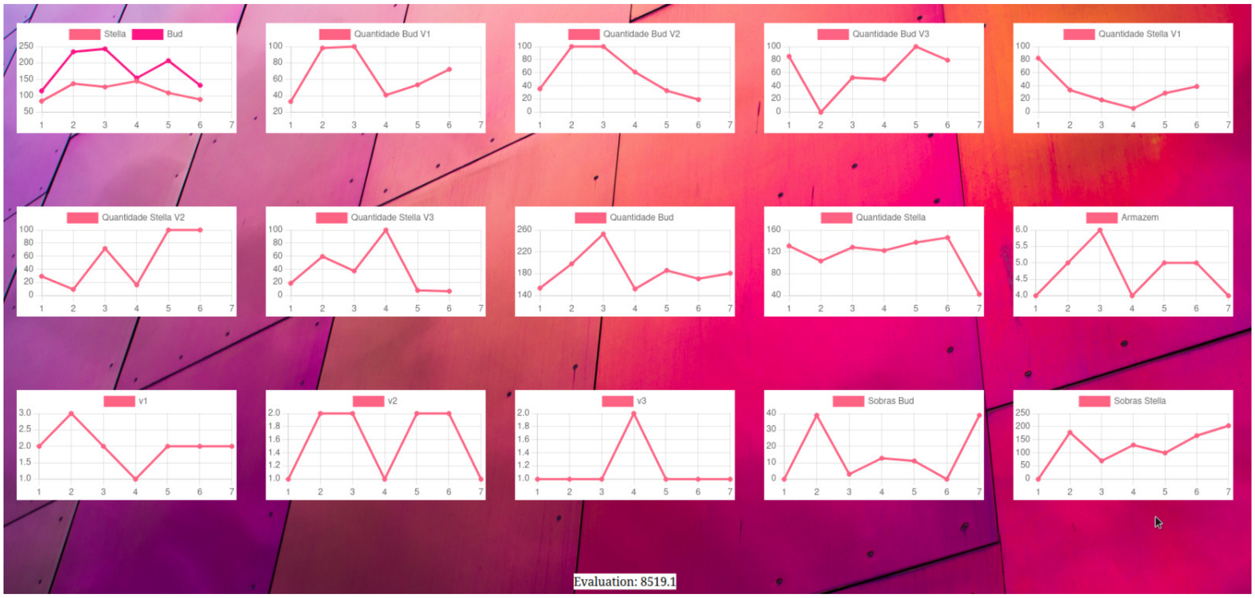

Figure 35: Multivariate Analysis

Comparison between Weeks 3 and 20:

Comparing the best results achieved by each algorithm in Weeks 3 and 20, the following observations can be made:

•Hill Climbing: In terms of production predictions, Hill Climbing maintained the same predictions for Stella and Bud beer production in both weeks. However, the evaluation score was higher in Week 20 (8471.8) compared to Week 3 (3879.2), indicating improved optimization in the latter week.

•Monte Carlo: The Monte Carlo algorithm also maintained consistent production predictions for Stella and Bud beer in both weeks. However, the evaluation score was higher in Week 20 (8518.1) compared to Week 3 (3676.1), indicating better overall optimization in the latter week.

The comparison of optimal results achieved by the Hill Climbing and Monte Carlo algorithms in Weeks 3 and 20 highlights interesting insights. Both algorithms produced similar predictions for beer production in both weeks, but the evaluation scores demonstrated a preference for the Monte Carlo algorithm. It achieved slightly higher scores in both weeks, indicating better overall performance in optimizing production levels. Further analysis and experimentation are recommended to understand the strengths and weaknesses of these algorithms for beer production optimization across different time periods and evaluate their suitability for real-world brewery operations.

6 Development System Demonstration

Where is the link for the video of the demonstration:

•https://youtu.be/-FC5HtxhLC8

7 Conclusions

With this, we conclude the document related to the TIAPOSE project report, on the topic Support System for the Production and Distribution of Beverages.

With this work we managed to create a way of working where all team members collaborated to carry out the work in the most efficient and effective way.

Also thanks to this work we were able to improve the forecasting and optimization techniques that we already had knowledge of and we were also able to learn new techniques.

8 Attachments

Github project link:

- https://github.com/Dong-Xuyong/TIAPOSE

Kaggle Notebook (Version 27):

- https://www.kaggle.com/code/dongxuyong/drinks-eda/notebook